

---

**AVOIMEN DATAN MAHDOLLISUUDET  
MATKAILUYRITYSTEN  
LIIKETOIMINNALLE**



Ammattikorkeakoulun opinnäytetyö

Tietojenkäsittelyn koulutusohjelma

Visamäki, kevät 2015

Joni Kukkamäki



## VISAMÄKI

Tietojenkäsittelyn koulutusohjelma  
Systeemityö

---

<b>Tekijä</b>	Joni Kukkamäki	<b>Vuosi</b> 2015
<b>Työn nimi</b>	Avoimen datan mahdollisuudet matkailuyritysten liiketoiminnalle.	

---

## TIIVISTELMÄ

Opinnäytetyön toimeksiantaja oli Hämeen ammattikorkeakoulun älykkäät palvelut -tutkimusyksikkö. Sen tehtävänä on kehittää digitaalisuuden käyttöä palveluliiketoiminnassa sekä tuottaa yhteistoimintaverkoston kanssa soveltavaa tutkimusta ja aluekehitysaktiviteetteja alueen tarpeisiin. Tämän vuoksi myös avoimen datan tuntemus on yksi merkittävistä tutkimusyksikön tarpeista.

Työn tarkoituksena oli selvittää avoimen datan mahdollisuuksia matkailuyritysten liiketoiminnan kehittämisessä. Opinnäytetyö toteutettiin myös niin, että se voisi toimia oppaana yrityksille, jotka haluavat tutustua laajemmin avoimen dataan.

Opinnäytetyössä käytettiin avoimen datan julkaisuista sekä internetlähteistä saatua teoretietoa. Tietoa oli saatavilla kohtuullisesti, mutta koska ilmiö on verrattain tuore, aiempia tutkimuksia ei ole julkaistu vielä kovin paljon, varsinkaan koskien yrityksiä avoimen datan hyödyntäjinä. Tutkimusosuu- den aineisto kerättiin matkailualan sekä avoimen datan asiantuntijoille suunnatuilla teemahaastatteluilla.

Tutkimuksessa selvisi, että matkailuyritykset eivät vielä hyödynnä digita- lisaation mukana tuomia mahdollisuuksia tehokkaasti. Dataa ei kerätä omista toiminnoista riittävän tehokkaasti eikä sitä hyödynnetä riittävän hyvin. Matkailuyrityksissä avoimen datan käyttö on vielä hyvin alkuteki- jöissään, vaikka sillä voitaisiin saavuttaa merkittäviä hyötyjä. Suurin ha- vaittu yksittäinen syy tälle on resurssipula, sillä matkailuyritykset ovat usein kooltaan hyvin pieniä, eikä yrittäjillä riitä resurssit muuhun kuin ydinliiketoiminnan ylläpitoon. Suomessa kuitenkin on olemassa kaikki edellytykset avoimen datan hyödyntämiselle myös liiketoiminnassa: avointa dataa on saatavilla ja sen käyttöä tukevia toimia on käynnistetty laajasti.

**Avainsanat** avoin data, tieto, matkailu, liiketoiminta

**Sivut** 38 s. + liitteet 1 s.

VISAMÄKI

Degree Programme in Business Information Technology  
System Development

---

<b>Author</b>	Joni Kukkamäki	<b>Year</b> 2015
<b>Subject of Bachelor's thesis</b>	Possibilities of open data for tourism business.	

---

## ABSTRACT

This Bachelor's thesis was commissioned by the research unit Smart Services of Häme University of Applied Sciences. The goal of the unit is to develop the use of digitalization in the service business and to produce a network of co-operation within applied research and regional development activities for the needs of the region. Due to this, the open data knowledge is one of the major needs of the research unit.

The purpose of this thesis was to study the potential of open data in the business development of tourism companies. Another aim of the thesis was to act as a guidebook for companies that want to explore the open data in more detail.

The theoretical information was gathered from the publications and internet sources of open data. As the phenomenon is relatively recent, not many previous studies have been published, especially regarding the companies as a user of the open data. To collect research material experts in the field of tourism and open data were interviewed.

The results of the thesis show that tourism enterprises do not yet take advantage of the opportunities offered by digitalization effectively. Data on their own operations is not collected effectively enough and it is not utilized sufficiently. In tourism business, the use of open data is still at a very early stage of development, even though it could be used to gain remarkable benefits. The main single reason found for this is the shortage of resources, because tourism companies are often very small in size, and entrepreneurs have resources only for maintaining their core business activities. In Finland, however, there are all the conditions to utilize open data also in business as open data is available and many kind of support activities have been launched extensively.

**Keywords** open data, information, tourism, business

**Pages** 38 p. + appendices 1 p.

---

## SANASTO

<b>API</b>	Application Programming Interface. Ohjelmointirajapinta.
<b>CSV</b>	Comma-separated values. Tekstitiedostomuoto, jossa tietoja erotellaan pilkkujen ja rivinvaihtojen avulla.
<b>ETL</b>	Extract – Transform – Load. Vaihe, jossa eri perusjärjestelmien tiedot luetaan, muutetaan vaadittavaan muotoon ja ladataan tietovarastoon.
<b>GeoRSS</b>	Metatietoa sisältävä maantieteellinen syöte.
<b>GPS</b>	Global Positioning System. Maailmanlaajuinen paikallistamisjärjestelmä.
<b>Javascript</b>	Web-ympäristössä käytettävä dynaaminen komentosarjakieli.
<b>JSON</b>	Javascript Object Notation. Avoimen standardin tiedostomuoto.
<b>KML</b>	Keyhole Markup Language. Merkintäkieli maantieteellisen tiedon esittämiseen.
<b>Metatieto</b>	Kuvailu- ja määrittelytieto.
<b>ODS</b>	OpenDocument-laskentataulukkotiedosto.
<b>OLAP</b>	Online analytical processing. Tapa analysoida tietoa moniulotteisesti.
<b>PDF</b>	Portable Document Format. Sähköisessä julkaisemisessa käytetty tiedostomuoto.
<b>XML</b>	Extensible Markup Language. Merkintäkielistandardi.

# SISÄLLYS

1	JOHDANTO.....	1
1.1	Tutkimuksen tavoitteet, tutkimuskysymykset ja rajaukset .....	2
1.2	Toimeksiantajan esittely.....	3
2	AVOIN DATA .....	4
2.1	Datan määrittely .....	4
2.1.1	Datatietoisuus .....	5
2.1.2	Tietoarkkitehtuuri .....	7
2.1.3	Tietovarastointi (Data Warehousing) .....	7
2.1.4	Business intelligence .....	8
2.2	Avoimen datan kriteerit.....	9
2.2.1	Maksuttomuus .....	9
2.2.2	Saatavuus ja löydettävyys.....	10
2.2.3	Uudelleenkäyttö ja -jakelu.....	10
2.2.4	Koneluettavuus ja formaatin avoimuus .....	10
2.2.5	Käyttöehtojen tasa-arvoisuus.....	11
2.2.6	Ymmärrettävyys .....	11
2.2.7	Viittaaminen ja eheys .....	11
2.3	Työkalut avoimen datan hyödyntämiseen.....	11
2.3.1	Ohjelmointirajapinnat .....	11
2.3.2	Tiedostoformaatit .....	12
2.3.3	Datakatalogit.....	13
2.3.4	Lisensointimenetelmät.....	13
2.3.5	Linkitetty avoin data .....	15
2.3.6	Big Data.....	16
2.4	Tiedon avaaminen ja julkaiseminen .....	16
2.4.1	Suunnittelu ja toteutus .....	17
2.4.2	Lainsäädäntö.....	17
2.4.3	Anonymisointi .....	18
2.5	Avoin data Suomessa .....	19
2.5.1	Tietoa merkittävimmistä tiedon avaajista.....	19
2.5.2	Datakatalogeja .....	20
2.5.3	Hankkeet ja organisaatiot .....	21
3	TUTKIMUSMENETELMÄT JA TUTKIMUKSEN TOTEUTUS .....	23
3.1	Tiedonkeruumenetelmät.....	23
3.2	Analysointimenetelmät.....	24
3.3	Haastateltujen asiantuntijoiden esitleminen .....	25
3.4	Tutkimuksen luotettavuus .....	25
4	TUTKIMUSTULOKSET .....	27
5	JOHTOPÄÄTÖKSET JA JATKOKEHITYSEHDOTUKSET .....	31
	LÄHTEET .....	34
	HAASTATTELUT .....	38

---

Liite Teemahaastatteluissa käytetty haastattelurunko

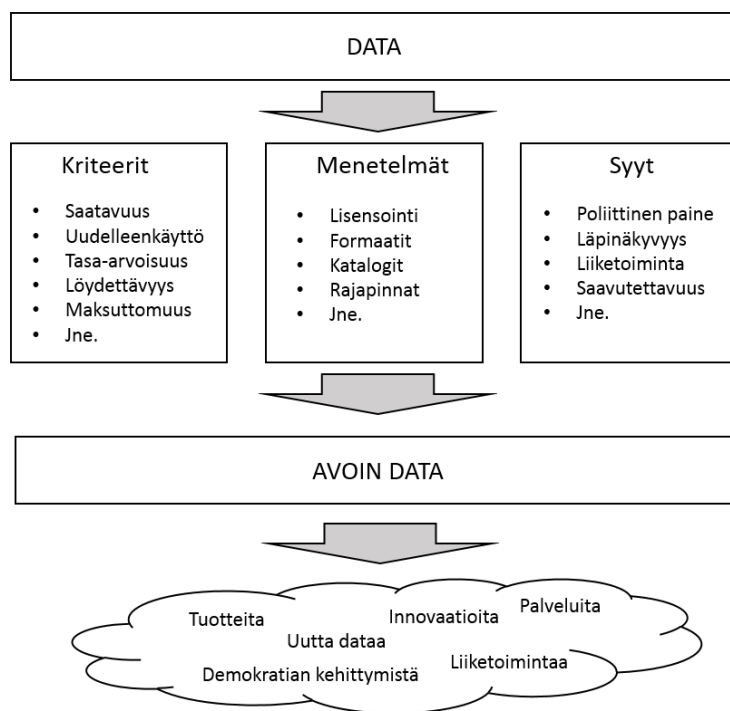
## 1 JOHDANTO

Avoim data on suhteellisen tuore ilmiö. Yleisesti sillä tarkoitetaan avoimesti ja ilmaiseksi jaettua, koneluettavassa muodossa olevaa dataa, jonka käyttöehdot mahdollistavat sen uudelleenkäytön. (Poikola, Kola & Hintikka 2010, 35–38.) Avoimelle datalle ei kuitenkaan ole olemassa yksiselitteistä määritelmää, ja se voidaan nähdä enemmänkin tapana toimia. Sen juuret ovat Yhdysvalloissa ja Isossa-Britanniassa, joissa myös poliittinen paine datavarantojen avaamista kohtaan on ollut suuri. Avoimen datan hyötyjä aletaan kuitenkin vasta nyt todella ymmärtää. Julkishallinnon organisaatiot ovat alkaneet avata datavarantojaan ja tietyt aktiiviset tahot ovat alkaneet hyödyntää tätä raaka-ainetta. Avoim data onkin yhdessä big datan kanssa yksi tämän hetken merkittävimmistä digitalisaation ilmiöistä. On jopa arvioitu, että avoimella datalla saavutettavat hyödyt voisivat olla suurempia kuin big datan kohdalla.

Avoimen datan ympärille ei kuitenkaan ole syntynyt sellaista innostusta kuin esimerkiksi internet of thingsin tai big datan ympärille. Tämä johtuu pääsääntöisesti siitä, että avoim data on yleisesti yksitoikkoista julkishallinnon dataa eikä sitä koeta mielenkiintoiseksi tai nähdä sen hyötyjä. (Peltola 2014.) Pääsääntöisesti avointa dataa onkin tähän asti käytetty lähinnä kansalaisaktiivisiin. Tämäkin tukee demokratiaa ja edistää palveluiden kehittymistä, mutta seuraavaksi tulisi alkaa hyödyntää näitä isoja datavarantoja ja etsiä niistä mahdollisuuksia myös liiketoimintakäyttöön.

Tähän asti dataa ovat siis avanneet lähinnä julkishallinnon organisaatiot. Myös yksityisten yritysten tulisi nyt alkaa kartoittaa mahdollisuuksia avata datavarantojaan ja miettiä mitä hyötyjä sillä voidaan saavuttaa. Usein hyöty yhdistetään liiketoiminnassa ainoastaan rahaan. Hyötyä voidaan saavuttaa kuitenkin myös esimerkiksi brändin imagon tai verkostojen toiminnan kehittymisenä, jotka taas lopulta lisäävät liikevaihdon kasvua. Tämä on vielä kartoittamatonta aluetta, ja nyt vaadittaisiinkin rohkeita ja innovatiivisia toimia myös yrityksiltä.

Avoim data itsessään ei kuitenkaan mahdollista automaattisesti merkittäviä muutoksia liiketoiminnalle kuten ei sen yläkäsite digitaalisuuskaan. Ne tulisi nähdä mahdollistavina alustoina, joiden avulla voidaan synnyttää uusia innovaatioita ja liiketoimintaa. Tämän opinnäytetyön tarkoituksena on selvittää, miten tätä ilmiötä voitaisiin hyödyntää tämän kaltaisessa innovatiivisessa kehitystyössä. Se ei niinkään ota kantaa siihen, minkälaisia uusia palveluita tai muutoksia nykyisiin palveluihin matkailuyritysten tulisi kehittää. Opinnäytetyö esittelee ilmiöön liittyviä käsitteitä, kriteereitä, teknologioita, menetelmiä ja syitä sekä datan avaamista tukevia tahoja (Kuva 1).



Kuva 1. Datan avauksen prosessi ja avoimen datan mahdollisuudet

Matkailualalla on merkittävä rooli Suomen kansantaloudelle. Kotimaiset ja ulkomaiset matkailijat kuluttivat Suomessa vuonna 2013 lähes 14,5 miljardia euroa. Matkailualan arvonlisäys oli vuonna 2013 lähes 2,5 % Suomen bruttokansantuotteesta ja sen on huomattavan iso osuus verrattuna moneen muuhun toimialaan. Maailmalla on arvioitu matkailun kasvavan 4 % vuosivauhtia. Matkailualalla on myös suuret työllistämisaikutukset: vuonna 2013 Suomessa oli noin 27 000 matkailualan yritystä, jotka työllistivät yhteensä 140 000 työntekijää. (Työ- ja elinkeinoministeriö 2015.)

### 1.1 Tutkimuksen tavoitteet, tutkimuskysymykset ja rajaukset

Tutkimuksen tavoitteena on selvittää matkailuyritysten mahdollisuuksia, halua ja kykyä hyödyntää avointa dataa liiketoimintansa kehittämisessä. Tavoitteena on myös, että tämä opinnäytetyö voisi toimia oppaan kaltaisena julkaisuna, jonka avulla yritykset voivat paremmin ymmärtää, mitä on avoin data ja mitkä sen mahdollisuudet ovat.

Opinnäytetyö etsii vastausta kysymykseen, miten matkailuyritykset voisivat hyödyntää avointa dataa liiketoiminnassaan. Tähän tutkimuskysymykseen liittyvät myös seuraavat tarkentavat tutkimuskysymykset: Mitä on avoin data? Miten avointa dataa tulisi käyttää? Mitä hyötyä omien datavarojen avaamisella voitaisiin saavuttaa? Mitä edellytyksiä avoimen datan käyttöön liittyy?

Opinnäytetyö on rajattu koskemaan ainoastaan matkailuyritysten mahdollisuuksia hyödyntää avointa dataa, joten se ei käsittele sitä, miten yksityishenkilöt voivat dataa käyttää. Datan avaajan roolissa kuitenkin käsitellään myös julkishallinnollisia organisaatioita. Tämän työn tarkoitus ei ole tarjo-



ta valmiita ratkaisuita vaan esitellä avoin data alustana ja mahdollisuutena luoda uusia innovaatioita, palveluita ja liiketoimintaa. Opinnäytetyössä datalla tarkoitetaan ainoastaan digitaalista aineistoa.

### 1.2 Toimeksiantajan esittely

Opinnäytetyön toimeksiantaja on Hämeen ammattikorkeakoulun älykkäät palvelut -tutkimusyksikkö. Se tuottaa yhteistoimintaverkoston kanssa soveltavaa tutkimusta ja aluekehitysaktiviteetteja alueen tarpeisiin. Yksikön toiminta kattaa digitaalisuuden ja palveluliiketoiminnan kehittämisen yli toimialarajojen, sillä usein samanlaiset ratkaisut toimivat soveltaen eri toimialoilla. (Hämeen ammattikorkeakoulu. n.d-d.)

Hämeen ammattikorkeakoulu eli HAMK on alueensa ainoa korkeakoulu. Vuodesta 2015 lähtien se toimii osakeyhtiönä, ja sen osakaskuntia ovat Forssa, Hattula, Hämeenlinna, Riihimäki, Tammela ja Valkeakoski. HAMK:n toiminta on järjestetty viiteen eri yksikköön, joiden tehtävänä on järjestää opetusta, tehdä tutkimusta, toimia yhteistyössä yritysten sekä yhteisöjen kanssa ja ylläpitää kansainvälistä toimintaa. Tutkimuksen ja koulutuksen yhteistoiminnalla edistetään palveluiden ja osaamisen saatavuutta. (Hämeen ammattikorkeakoulu. n.d-c.)

HAMK tarjoaa koulutusta seitsemällä eri kampuksella. Koulutusohjelmia on yhteensä noin 30. HAMKissa voi suorittaa amk-tutkinnon tai ylemmän amk-tutkinnon. Tutkintoja järjestetään sekä suomeksi että englanniksi. HAMK tarjoaa myös ammatillista opettajakoulutusta sekä täydennyskoulutuksia. (Hämeen ammattikorkeakoulu. n.d-a.)

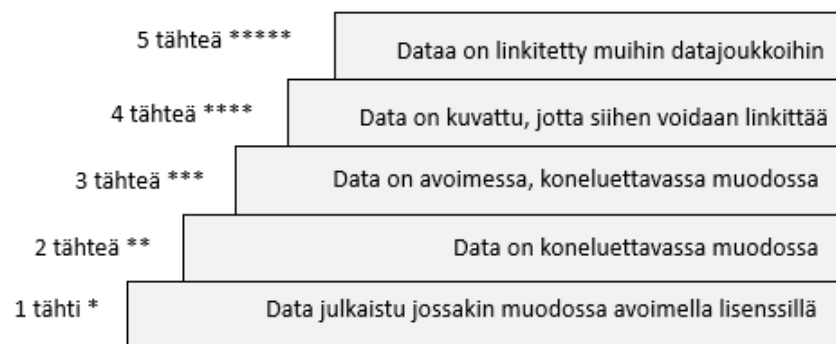
HAMK:n tutkimustoiminta on pääsääntöisesti soveltavaa tutkimusta ja se perustuu sen neljään tutkimusyksikköön: Ammatillinen osaaminen, biotalous, ohutlevykeskus ja älykkäät palvelut. Nämä tutkimusyksiköt edistävät aluevaikuttavuutta ja työelämälähtöistä tutkimusta. (Hämeen ammattikorkeakoulu. n.d-b.)

## 2 AVOIN DATA

Nykypäivänä kaikenlainen kerättävissä oleva data on yrityksille erittäin tärkeää raaka-ainetta. Datan käyttö liiketoiminnassa ei ole enää vain IT-yritysten ja korporaatioden toimintaa, vaan jokaisen yrityksen tulisi koostaa tai toimialasta huolimatta kartoittaa mahdollisuutensa hyödyntää dataa liiketoiminnan kehittämisessä ja lisätä omaa ymmärrystään.

Dataa on mahdollista kerätä kaikkialta, vaikka sitä ei aina osata tiedostaa. Sitä on saatavilla yksinkertaisimmillaan omista jokapäiväisistä järjestelmistä kuten maksuliikenteestä, kotisivuilta ja sähköposteista. Asiakashallinta- ja toiminnanohjausjärjestelmät tuottavat uutta dataa jatkuvasti. Sitä on saatavilla myös oman yrityksen ulkopuolelta maksua vastaan ja myös täysin ilmaiseksi. Julkishallinto ja jossain määrin myös yritykset ovat alkaneet avata datavarantojaan kaikkien saataville ja käytettäväksi. Tällöin puhutaan avoimesta datasta. (Lohtander 2013, 3-5; Open Knowledge Finland 2014.)

Avoimelle datalle on olemassa monenlaisia kriteeristöjä, ja osa näistä kriteereistä esitellään tässä opinnäytetyössä. Olennaisinta kuitenkin on, että datan on oltava kaikkien saatavilla, ilmaiseksi ja koneluettavassa muodossa. Datan laatua voidaan luokitella esimerkiksi viisiportaisen menetelmän avulla, joka on esitetty alla olevassa kuvassa (Kuva 2). Kuten siitä voi havaita, avoimen datan määritelmä on mahdollista saavuttaa jo pienillä toimenpiteillä, mutta laadukkaamman datan määrittäminen edellyttää haastavampia toimia.

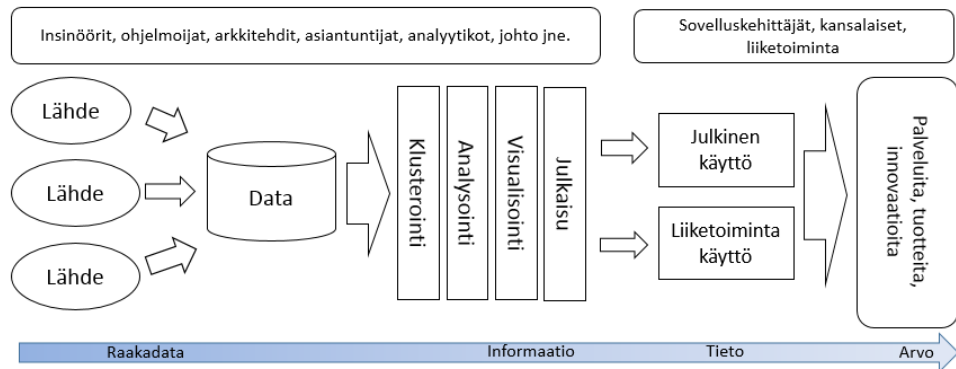


Kuva 2. Avoimen datan laadullinen luokittelu (Lohtander 2013, 38).

### 2.1 Datan määrittely

Yleisesti englanninkielinen termi data käännetään suomen kielessä tiedoksi (vrt. database = tietokanta). Tämä saattaa kuitenkin aiheuttaa väärinkäsityksiä, sillä datalla tarkoitetaan raaka-ainetta, josta voidaan muodostaa merkityksellisempää informaatiota ja sitä edelleen jalostamalla tietoa. Tieto on suomen kielessä monimerkityksellinen sana, ja tämän vuoksi on viisaampaa puhua datasta käsiteltäessä digitaalisessa muodossa olevaa raakamateriaalia. Suomessa on myös puhuttu tietovarannoista, joka olisikin paremmin kuvaava suomennos sanalle data. (Poikola ym. 2010, 34.)

Data on siis digitaalisessa muodossa olevaa raaka-ainetta. Se voi olla ääntä, videota, kuvaa, tekstiä, tietokantoja ja niin edelleen. Yleisesti sillä ei kuitenkaan ole arvoa itsessään, vaan se vaatii prosessointia, jotta sitä voidaan hyödyntää arvontuotossa.

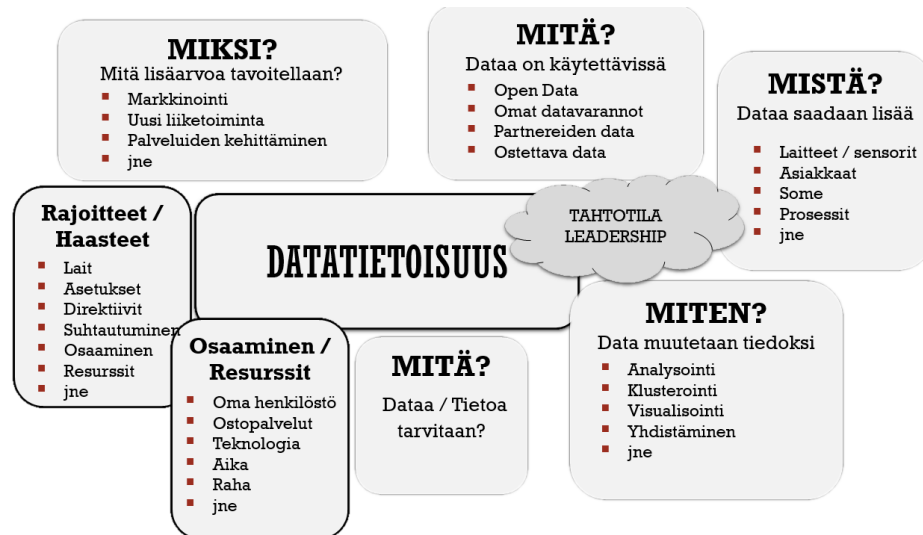


Kuva 3. Datin hyödyntäminen arvontuotossa

Dataa voidaan kerryttää useista erilaisista lähteistä kuten käyttäjiltä, antureista, olemassa olevista varannoista, transaktioista ja sosiaalisesta mediasta. Kaikki kerätty data on varastoitava, ja sen jälkeen sitä voidaan alkaa käsitellä. Dataa analysoitaessa siitä muodostuu informaatiota, jota käsittelemällä voidaan tuottaa tietoa liiketoiminnan tueksi tai julkaistavaksi (Kuva 3). Tätä tietoa hyödyntämällä voidaan luoda uusia palveluita ja tuotteita. (Liikenne- ja viestintäministeriö. 2014, 9.)

### 2.1.1 Datatietoisuus

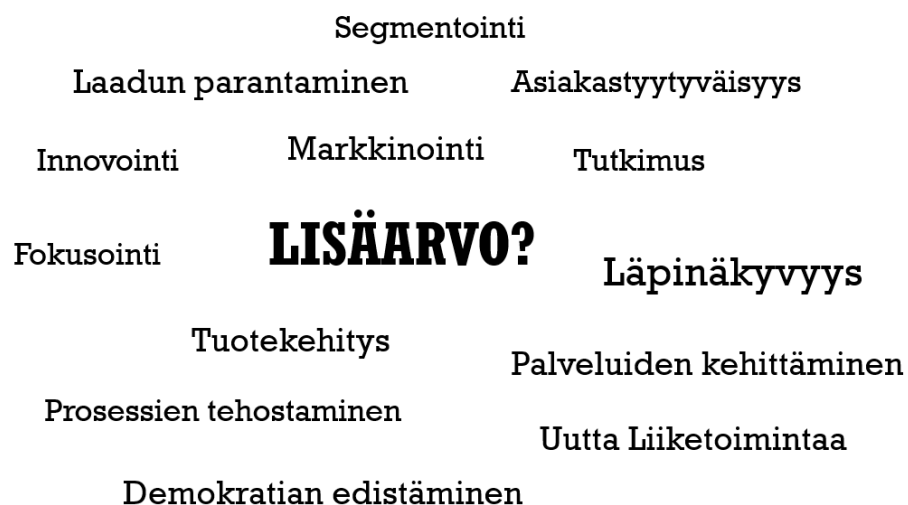
Datatietoisuudella viitataan esimerkiksi yrityksen, hallinnon tai tutkimuksen kykyyn ymmärtää datassa piilevä potentiaali, mutta myös riskit. Keräämällä tarpeeksi laadukasta dataa ja analysoimalla sitä oikein voidaan saavuttaa merkittävää lisäarvoa. Ensisijaista on osata etsiä oikeita käyttökohteita ja yhdistää siihen oikea käytettävissä oleva data. Datatietoisuus on siis myös mitä enimmäisessä määrin sidoksissa muutosjohtajuuteen, ja sen tulee näkyä jokapäiväisessä toiminnassa. (Liikenne- ja viestintäministeriö. 2014, 14–15.)



Kuva 4. Datatietoisuuden lisääminen kysymysten avulla

Datatietoisuutta on helppo kasvattaa esittämällä erilaisia kysymyksiä (Kuva 4). Yritysten normaali tapa on hyödyntää ainoastaan omia sisäisiä datavarantoja ja niitäkin rajallisesti. Laajemmalla näkökulmalla voidaan oman organisaation ulkopuolelta löytää helposti saatavilla olevaa dataa, jota pystytään hyödyntämään omassa liiketoiminnassa. Miettimällä datankäsittelyprosessia uudelleen voidaan päästä tehokkaampiin toimintatapoihin ja havaita uusia käyttötarkoituksia. (Liikenne- ja viestintäministeriö. 2014, 14–15.)

Tärkein kysymys onkin, miksi datankäyttöä tulisi tehostaa ja mitä lisäarvoa sillä tavoitellaan. Vastaava kysymys toimii myös kaikissa muissa liiketoiminnan prosesseissa, ja tämän vuoksi samassa yhteydessä tulisi asiaa miettiä myös datan käytön näkökulmasta. Tällä voidaan ohjata myös resursseja: mikäli toimenpiteet nähdään kriittisinä liiketoiminnan kannalta, tulee niitä varten myös varata riittävästi rahaa ja työaika.

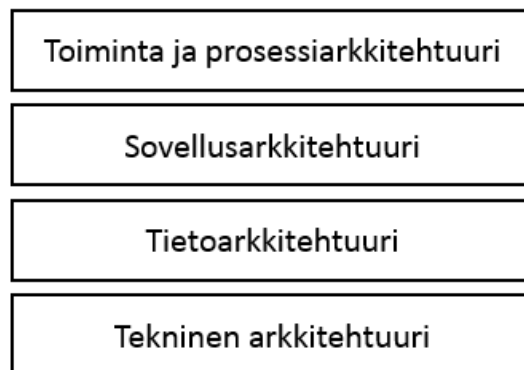


Kuva 5. Tehokkaalla datan käytöllä voidaan saavuttaa monenlaista lisäarvoa

### 2.1.2 Tietoarkkitehtuuri

Tietoarkkitehtuurilla tarkoitetaan yrityksen tai organisaation tietorakenteiden kokonaissuunnittelua. Se on hyvin laaja kuvaus kokonaisuudesta eikä pidä sisällään yksityiskohtaista tietoa järjestelmistä tai datavarannoista. Suunnittelulla on erittäin suuri merkitys toimivan kokonaisuuden luomisessa, ja siihen tarvitaan tietoarkkitehtuuria. Se parantaa tietojen välistä semantiikkaa ja sen ymmärtämistä. Kun on tiedossa mitä tietoa on ja missä, pystytään myös kehittämään tiedon integraatiota ja käytettävyyttä. Tiedon siiloutuminen järjestelmäkohtaiseksi tai organisaatioyksikkökohtaiseksi heikentää tiedon saatavuutta ja tehokkuutta. Usein esimerkiksi markkinointiosastolla saattaa olla omat tietovarastot mutta muilla osastoilla ei ole pääsyä näihin tai niiden olemassaolosta ei edes tiedetä. (Hovi 2009, 12–14.)

Laajan kokonaisarkkitehtuurin perusteella voidaan alkaa luoda yksityiskohtaisempia piirustuksia ja mallinnuksia, joita eri ammattilaiset tarvitsevat rakentaessaan ja suunnitellessaan laitteistoja, järjestelmiä ja ohjelmistoja. Tietoarkkitehtuurin suunnittelu taas vastaavasti perustuu yrityksen kokonaisvaltaisempaan arkkitehtuuriin. Näin onkin tärkeää, että ylimmältä suunnittelun tasolta aina yksityiskohtaisimpiin teknisiin toteutussuunnitelmiin toiminnot ovat samassa linjassa. (Hovi 2009, 12–14.)



Kuva 6. Arkkitehtuurirakenne (Hovi 2009, 14)

### 2.1.3 Tietovarastointi (Data Warehousing)

Kun päätöksenteossa tarvittavaa tietoa kerätään useista eri lähteistä, kuten omista operatiivisista järjestelmistä, ulkoisista järjestelmistä ja esimerkiksi eri avoimen datan lähteistä, tiedot on hyvä varastoida yhtenevässä muodossa yhteen tietovarastoon. Tästä sijainnista voidaan tämän jälkeen suorittaa Business intelligence -ohjelmistoilla erilaisia analyyseja ja raportteja. Tietovaraston tarkoituksena on mahdollistaa mahdollisimman helppo ja nopea tiedon saatavuus.

ETL-vaiheessa (extract - transform - load) eri perusjärjestelmien tiedot luetaan ja muutetaan vaadittavaan muotoon ja ladataan tietovarastoon. ETL-vaiheessa suoritettava yhdenmukaistaminen on tärkeää, sillä eri operatiivisten tietojärjestelmien omissa tietokannoissa merkitykset saattavat erota toisistaan. Kahteen eri tietokantaan voidaan kerätä kyllä/ei-tyyppisiä

vastauksia, ja koodata ne käyttäen numeroita. Tietokannassa 1 merkintäpa voi olla Kyllä = 0 Ei = 1, kun taas tietokannassa 2 se voi olla päinvastoin, Kyllä = 1 Ei = 0. Jos tietoja ei yhdenmukaisteta ETL-vaiheessa, tietovarastoon tallentuu virheellistä tietoa. (Hovi, Hervonen & Koistinen 2009, 14–15.)

ETL-vaiheessa voidaan myös jalostaa tietoja jo valmiiksi paremmin käytettävään muotoon. Esimerkiksi henkilötunnuksesta voidaan jo tässä vaiheessa erotella henkilön ikä omaan sarakkeeseen, mikäli tieto koetaan tarpeelliseksi. Usein tässä vaiheessa myös tehdään erilaisia tarkistuksia, joilla varmennetaan tietojen oikeellisuutta. Viimeisenä toimenpiteenä ETL-vaiheessa on tietojen lataaminen tietovarastoon. (Hovi ym. 2009, 14–15.)

ETL-vaihe voidaan toteuttaa aikataulutetusti, yleensä kerran päivässä. Tämä vaihe kannattaa toteuttaa ajankohtana, jolloin tarve tiedolle on mahdollisimman pieni, eli usein yöaikaan. Kun tiedot on siirretty tietovarastoon, sitä käytetään ainoastaan tietojen lukemiseen. Tarvittavat muutokset tehdään operatiivisten järjestelmien tietokantoihin. Näin varmistetaan tietojen ajantasaisuus. (Hovi ym. 2009, 14–15.)

Tietovarastointi tuo mukanaan useita eri hyötyjä kuten yhdenmukaisuutta toimintatapoihin, riippumattomuutta järjestelmistä ja nopeat ja helpot kyselyt. Tietovarastoja myös pidetään ”organisaatioiden muistina”, sillä niihin tallennettavat, jatkuvasti kasvavat tietomäärät pitävät sisällään historiatietoa, jota voidaan hyödyntää erilaisissa aika-analyyseissa. Koska tiedot on tallennettu tietovarastoon yhtenevässä muodossa, aineisto on helposti vertailtavissa myös pidempiä ajanjaksoja taaksepäin. (Hovi ym. 2009, 16.)

### 2.1.4 Business intelligence

Business intelligence (BI) on käsite, jolla tarkoitetaan kerätyn tiedon hyödyntämistä liiketoiminnan päätöksenteossa tuottamalla siitä erilaisia analyyseja ja raportteja, joita voidaan jakaa myös eteenpäin. Termiä on yritetty suomentaa usein eri tavoin, kuten liiketoimintatiedon hallintana. Termi business intelligence on kuitenkin yleiskäsitteenä niin vakiintunut, että sen käyttö on viisasta myös Suomen kielessä. Business intelligencen tavoitteena on, että tieto on jalostettu jo valmiiksi sellaiseen muotoon, että yrityksen henkilöstö pystyy hyödyntämään sitä ilman tietoteknistä osaamista käyttäen apuna siihen tarkoitettuja työkaluja. Parhaimmillaan BI-työkalut ovat siis erittäin yksinkertaisia ja helppokäyttöisiä käyttöliittymiä. Näistä tunnetuin lienee Microsoftin Excel ja sen pivot-työkalu, mutta Microsoftin lisäksi markkinoilla on myös muiden isojen yritysten tuotteita, esimerkiksi SAP, Oracle ja IBM. (Hovi ym. 2009, 74–77.)

Business intelligence -ratkaisulla voidaan saavuttaa useita merkittäviä hyötyjä. Se nopeuttaa päätöksentekoa ja tarjoaa luotettavaa tietoa vastineeksi mutu-tuntumalle. BI-ratkaisut edistävät myös organisaation henkilöstön omatoimisuutta, sillä tieto on helposti heidän saatavillaan eikä teknistä osaamista tarvita. BI-ratkaisuille on myös todellinen tarve, sillä yritysten ja organisaatioiden päätöksentekonopeus on kasvanut samalla kun

käytettävissä olevan informaation määrä on lisääntynyt huomasti. Kun aiemmin päätöksiä saatettiin tehdä tiedon perusteella, joka päivittyi vuosittain tai kuukausittain, nykyaikana päätöksenteon tukena olevat tiedot saatavat päivittyä päivittäin tai lähes reaaliajassa. (Hovi ym. 2009, 74–76.)

Business intelligence tarkoittaa numeroiden murskausta. Sen lähestymistapa on siis kvantitatiivinen, sillä data on yleensä numeraalista ja tallennettu strukturoidussa muodossa. BI:n yhteydessä puhutaan usein OLAP-analysointimenetelmästä, joka tarkoittaa tiedon moniulotteisuutta. Tiedolla on numeraalisesti mitattavia arvoja, mutta sitä voidaan tarkastella monesta eri näkökulmasta, kuten alueellisesti, ajanjaksollisesti tai asiakassegmenttien kautta. Tällä tarkastelutavalla samasta tiedosta saadaan irti paljon erilaista informaatiota päätöksenteon tueksi. (Hovi ym. 2009, 77–82.)

Nykypäivänä yritys voi kerätä myös monimuotoisempaa, strukturoimatonta dataa ulkoisista lähteistä. Tällaisia lähteitä voi olla esimerkiksi kilpailijoiden tai asiakkaiden julkaisemat kirjoitukset, uutiset ja sosiaalinen media. Tämä on kvalitatiivisempi näkökulma business intelligenceen nähden ja tällöin puhutaankin usein market intelligencesta, joka vaatii erilaista lähestymistapaa ja teknologiaa. (Hovi ym. 2009, 78–79.)

### 2.2 Avoimen datan kriteerit

Datan avoimuus ei ole yksiselitteinen asia. Avoimuudelle on monia eri määritelmiä ja kriteereitä. Yksittäistenkin tunnuspiirteiden täytyminen aineiston kohdalla on hyvä askel matkalla kohti avoimuutta, mutta varsinaiseen avoimen datan määritelmään usein liitetään useampien kriteereiden täytyminen. Toisaalta taas harvoin kaikki kriteerit täyttyvät, joten voidaan todeta, että mitä useampi sen parempi. Pääsääntönä voidaan pitää, että avoimen datan tulisi olla kaikkien saatavilla, ilmaiseksi ja koneluettavassa muodossa, eivätkä sen käyttöehdot tai tekniset ratkaisut saa asettaa merkittäviä hidasteita tai esteitä datan käytölle. (Poikola ym. 2010, 34.)

Yleisesti avoimeen dataan liitetyt kriteerit perustuvat Open Definition -määritelmään, joka määrittelee avoimuuden tunnuspiirteitä. Sen on laatinut Open Knowledge, joka on maailmanlaajuinen voittoa tavoittelematon järjestö. Sen tehtävänä on edistää avoimen datan käyttöä ja lisätä sen tunnettua. (Open Knowledge. n.d.)

Open Knowledge Finland on Open Knowledgen Suomessa toimiva paikallinen jaosto. Sen tehtävänä on edistää avoimuutta alueellisesti Suomessa ja se on esimerkiksi suomentanut Open definition -määritelmän. (Open Knowledge Finland. n.d-b.)

#### 2.2.1 Maksuttomuus

Avoimen datan tulisi olla saatavissa täysin ilmaiseksi tai korkeintaan hyvin kohtuullista korvausta vastaan, jolla voidaan kattaa lähinnä jakeluun ja toimittamiseen liittyviä kuluja. Datavarantojen avaaminen on yleisesti prosessi, joka aiheuttaa suuriakin kuluja, joita ei tulisi periä datan käyttäjältä.

Maksullisuuteen liittyy usein sopimuksia ja byrokratiaa, jotka saattavat muodostaa rajoitteita datan käytölle. Tästä syystä täysin ilmaiseksi saatavilla oleva data on paras ratkaisu. (Poikola ym. 2010, 36–37.)

### 2.2.2 Saatavuus ja löydettävyys

Aineiston tulee olla täysin vapaasti saatavilla internetissä ja löydettävissä sieltä. Sen sijainti on yleisesti tiedossa, ja sen pystyy löytämään sekä ihminen että hakukoneet. Tämä koskee myös aineiston lisensointitietoja. Löydettävyyttä voidaan lisätä tiedottamalla siitä ja siihen kohdistuvista muutoksista aktiivisille loppukäyttäjille, lisäämällä riittävät aineiston kuvailutiedot ja linkittämällä aineistoa muuhun aineistoon. Usein myös oikean julkaisualustan, kuten datakatalogin valinta, edistää löydettävyyttä. (Poikola ym. 2010, 35.)

Aineisto tulisi myös olla ladattavissa kokonaisuutena eikä vain tiettyjä osia siitä. Usein tätä rajoitetaan tarjoamalla ainoastaan kyselyrajapintoja, jotka mahdollistavat pääsyn ainoastaan osaan aineistosta. Aineiston tarjoaminen kokonaisuudessaan mahdollistaa myös todellisen aineiston uudelleenkäytön ja -jakelun. (Poikola ym. 2010, 35.)

### 2.2.3 Uudelleenkäyttö ja -jakelu

Data tulee tarjota kaikkien käyttäjien saataville ja lisensoinnin tulee tarjota mahdollisuus uudelleenkäyttää ja -jaella aineistoa laillisesti ja vapaasti missä tarkoituksessa tahansa. Aineiston tarjoajan tulee siis myös luopua osasta tekijänoikeuksistaan. Käyttöehdot tulee esittää materiaalin yhteydessä selkeästi. Esimerkiksi Creative Commons -käyttöehdot tarjoavat toimivan ja valmiin ratkaisun. Tietoa avattaessa tulee kuitenkin huomioda, ettei julkaista materiaalia, jonka kohdalla saatettaisiin rikkoa eri lakeja tai asetuksia, kuten esimerkiksi yksityisyyden suoja tai muita salassapitomääräyksiä. (Poikola ym. 2010, 36.)

### 2.2.4 Koneluettavuus ja formaatin avoimuus

Tietojen avaaminen esimerkiksi PDF-formaatissa on ajatuksena hieno toimenpide ja hyvä alku, mutta se ei täytä vielä puhtaimman avoimen datan kriteereitä, sillä aineiston tulee olla siinä muodossa, että tietokoneet pystyvät käsittelemään sen suoraan ilman isoja työvaiheita välissä. Esimerkkinä mainittu PDF on tiedostomuoto, joka on helppolukuinen ihmisille, mutta koneille vaikea käsitellä. Siksi tulisikin käyttää muita avoimeen datan julkaisuun paremmin soveltuvia formaatteja. (Poikola ym. 2010, 36.)

Myös formaatin tulisi olla avoin, eli sen käyttöön ei saa liittyä rahallisia tai muita esteitä, kuten riippuvuutta tiettyjen yritysten valmistamista ohjelmistoista. Aineiston julkaisussa käytettyjen käyttöehtojen tulee myös olla saatavissa koneluettavassa muodossa. Sama tieto voidaan myös julkaista useammassa eri formaatissa. (Poikola ym. 2010, 36.)



### 2.2.5 Käyttöehtojen tasa-arvoisuus

Kuten saatavuuden kohdalla määritellään, datan tulee olla vapaasti saatavilla. Tätä ei voida siis myöskään rajoittaa kohdentamalla aineistoa vain tietyille käyttäjille ja ryhmille esimerkiksi rekisteröitymisten tai suljettujen sivustojen avulla. Tasa-arvoisuus siis tarjoaa kunnille, yhteisöille, yksityishenkilöille, yrityksille ja puolueille yhdenmukaiset oikeudet käyttää dataa omiin tarkoituksiin, olivat motiivit käytölle sitten kaupallisia, uskonnollisia tai poliittisia. (Poikola ym. 2010, 35.)

### 2.2.6 Ymmärrettävyys

Kuten kaiken datan kohdalla, myös avoin data kannattaa kuvailla ja dokumentoida selkeästi ja kattavasti. Dokumentointi lisää datan ymmärrettävyyttä ja sitä kautta käytettävyyttä. Käyttäjille voidaan antaa myös käyttötapaesimerkkejä. Metatietojen lisääminen aineistoon on myös tehokas keino lisätä aineiston ymmärrettävyyttä. (Poikola ym. 2010, 35.)

Dokumentointi on usein työläs prosessi, varsinkin jos sitä ei ole tehty samassa yhteydessä aineiston kanssa, vaan se on luotava jälkikäteen erillisenä projektina. Esimerkiksi ohjelmistokehityksessä dokumentointia laiminlyödään pahimmillaan erittäin räikeästi, mikä saattaa aiheuttaa paljon lisätyötä myöhemmissä vaiheissa kuten päivityksien yhteydessä. Jonkinasteinen dokumentointi kannattaakin luoda heti alkuvaiheessa, ja sitä voidaan laajentaa ajan kuluessa.

### 2.2.7 Viittaaminen ja eheys

Käyttöehtoja määriteltäessä lisenssi voi vaatia viittaamista alkuperäiseen aineistoon, kun sitä jaellaan tai käytetään uudelleen. Avoimen datan periaatteiden mukaisesti tämä ei kuitenkaan saa aiheuttaa kohtuutonta lisätyötä tai estää aineiston käyttöä. Aineiston yhteydessä tulisikin antaa tiedot joihin viitata. Lisenssi voi myös vaatia, että uusi aineisto nimetään uudestaan version mukaisesti alkuperäiseen aineistoon nähden. (Open Knowledge Finland. n.d-a.)

## 2.3 Työkalut avoimen datan hyödyntämiseen

Avoin data on enemmänkin tapa toimia eikä joukko teknisiä ratkaisuja. Avoimen datan käyttämisen ja datan avaamisen onnistumiseen vaikuttaa kuitenkin enemmän tahto eikä se, onko käytössä kalliita ohjelmistoja ja laitteita. Avoimen datan käsittelyyn ja julkaisuun liittyy kuitenkin tiettyjä oleellisia teknisiä toimenpiteitä, joista on hyvä olla tietty perusymmärrys, ennen kuin avaamista aletaan suunnittelemaan.

### 2.3.1 Ohjelmointirajapinnat

Ohjelmointirajapinta eli API (Application Programming Interface) tarkoittaa verkossa toimivaan sovellukseen rakennettua mahdollisuutta tehdä

sieltä erilaisia kutsuja muiden ohjelmien kautta. Tällainen kutsu on esimerkiksi kysely tietyistä tiedoista, johon vastauksena saadaan sitä koskeva data. Rajapintojen kautta on myös mahdollista syöttää tietoja järjestelmään, eli sen toiminta voi olla kaksisuuntaista. Tämä mahdollistaa ohjelmien integroinnin sekä omien sovellusten ja laajennusten luomisen. Rajapinnat voivat olla pelkkiä datarajapintoja, jolloin niiden välityksellä liikkuu vain dataa tai ne voivat olla toiminnallisia, jolloin niiden avulla pystytään toteuttamaan erilaisia toimintoja ja prosessoimaan dataa kohti informaatiota jo kyselyvaiheessa. (Kivekäs 2014; Cennoapp 2014.)

Tärkeä osa rajapintoja on niiden dokumentaatio. Ilman niitä rajapintojen käyttö on lähes mahdotonta. Dokumentaatioissa on kuvattava ainakin, mitä kutsuja on käytettävissä ja mitä nämä kutsut palauttavat. Hyvässä dokumentaatioissa on myös kuvattu rajapinnan toiminta yleisesti, mihin se on tarkoitettu ja mihin sitä voidaan käyttää. Myös esimerkkikoodit helpottavat käyttäjää uuteen rajapintaan tutustuttaessa (Gruenbaum 2010; Cennoapp 2014).

Rajapintojen käyttöä voidaan hallita ja rajoittaa API-avaimien avulla. Avaimen saannin ehtona voikin olla rekisteröityminen, maksaminen siitä tai yhteistyö rajapinnan toteuttajan kanssa. Rajapinnat voivat olla myös täysin suljettuja ja yrityksen omien sovellusten käyttöön tarkoitettuja, mutta niitä voidaan julkaista myös täysin avoimena.

Puhuttaessa avoimista rajapinnoista, on olemassa kriteerejä joiden on täyttyvä. Rajapinnan käytön tulee olla kaikille avointa mihin tarkoitukseen tahansa. Sen tulee olla dokumentoitu ja tämän dokumentaation tulee olla saatavilla vapaasti verkosta. Järjestelmän tiedot on kuvattava riittävällä tarkkuudella, jotta käyttöönotto voidaan tehdä. Rajapinta on myös voitava ottaa käyttöön vapaasti milloin tahansa. Mikäli se vaatii esimerkiksi rekisteröitymisen, on sen oltava automaattinen. Vaikka rajapinta olisi avoin, ei sen kautta saatavan datan ole oltava avointa vaikka usein näin onkin. Voikin olla, että avoin rajapinta tarjoaa kaikkien saataville vain osan datasta tai ei lainkaan. Vaatimuksena kuitenkin on, että vähintään testiaineisto on oltava saatavilla. (Kivekäs 2014.)

### 2.3.2 Tiedostoformaattit

Avoimen datan kriteereissä määritellään, että myös käytettävien tiedostoformaattien tulisi olla avoimia ja riippumattomia esimerkiksi kaupallisista sovelluksista. Lisäksi formaatin tulisi olla koneluetettava. Excel-taulukot ovat useille käyttäjille tuttu formaatti, mutta se on kaupallisen yrityksen (Microsoft) luoma ja on osittain riippuvainen heidän ohjelmistoistaan. Nykypäivänä on kuitenkin olemassa myös useita avoimia toimisto-ohjelmistoja, jotka pystyvät vaivatta käsittelemään Excelin luomia tiedostoja, joten onkin parempi avata tiedot siinä formaatissa kuin jättää kokonaan avaamatta.

Muita yleisiä avoimen datan tiedostoformaatteja ovat XML, CSV ja JSON. Extensible Markup Language eli XML on yleiskäyttöinen merkin-täkielistandardi. Siitä on johdettu useita eri käyttötarkoituksia varten luo-

tuja erikoistuneita merkintäkieliä, kuten KML ja GeoRSS, jotka on tarkoitettu paikkatietojen merkitsemiseen. Myös Excel-tiedostojen kaltainen, OpenDocument standardiin perustuva ODS tiedostomuoto perustuu XML:ään. CSV (Comma separated values) on tiedostomuoto, joka on avattavissa taulukkolaskentaohjelmilla ja jonka tiedot on eroteltu toisistaan pilkulla. JSON (Javascript Object Notation) on tekstipohjainen datansiirtoformaatti. Nimestään huolimatta se ei ole riippuvainen Javascriptistä. (Poikola ym. 2010, 64.)

### 2.3.3 Datakatalogit

Datakatalogi on luettelo tai eräänlainen hakemisto saatavilla olevasta avoimesta datasta. Sen perimmäinen tarkoitus on tarjota mahdollisuus löytää tarvitsemansa data yhdestä paikasta. Omaa dataa avattaessa on tärkeää ilmoittaa sen olemassaolosta olennaisille katalogeille, jotta käyttäjät voivat löytää tiedon.

Datakatalogit yhdistelevät metatietoa useista saatavilla olevista avoimista aineistoista. Niitä on tarjolla monella tasolla, ja ne voidaan luokitella esimerkiksi kansainvälisiin, kansallisiin, seudullisiin tai kunnallisiin. Lisäksi on olemassa yksityisten ylläpitämiä katalogeja, joilla aktiiviset käyttäjät ovat ryhtyneet paikkaamaan julkisten organisaatioiden puutteita. Yksityisiä katalogeja on syntynyt myös muista syistä, kuten halusta ylläpitää viranomaisista riippumatonta, omaa yhteisöllistä katalogia. (Poikola ym. 2010, 75–77.)

### 2.3.4 Lisensointimenetelmät

Avoimen dataan yhdistetyistä lisensointimenetelmistä tunnetuin on Creative Commons. Sen uusin versio on CC 4.0. Creative Commons on eikaupallinen organisaatio, jonka tarkoituksena on mahdollistaa tietojen jakaminen maksuttomien ja helppokäyttöisten lisenssien avulla. Creative Commons -lisenssit eivät korvaa tekijänoikeuksia, mutta ne mahdollistavat perinteisten ”kaikki oikeudet pidätetään” -ehtojen muokkaamisen sellaisiksi, että käyttäjällä on oikeus käyttää ja muokata teoksia. (Creative Commons. n.d-a.)

Halutun Creative Commons -lisenssin valitseminen perustuu muutamaankin yksinkertaiseen kysymykseen:

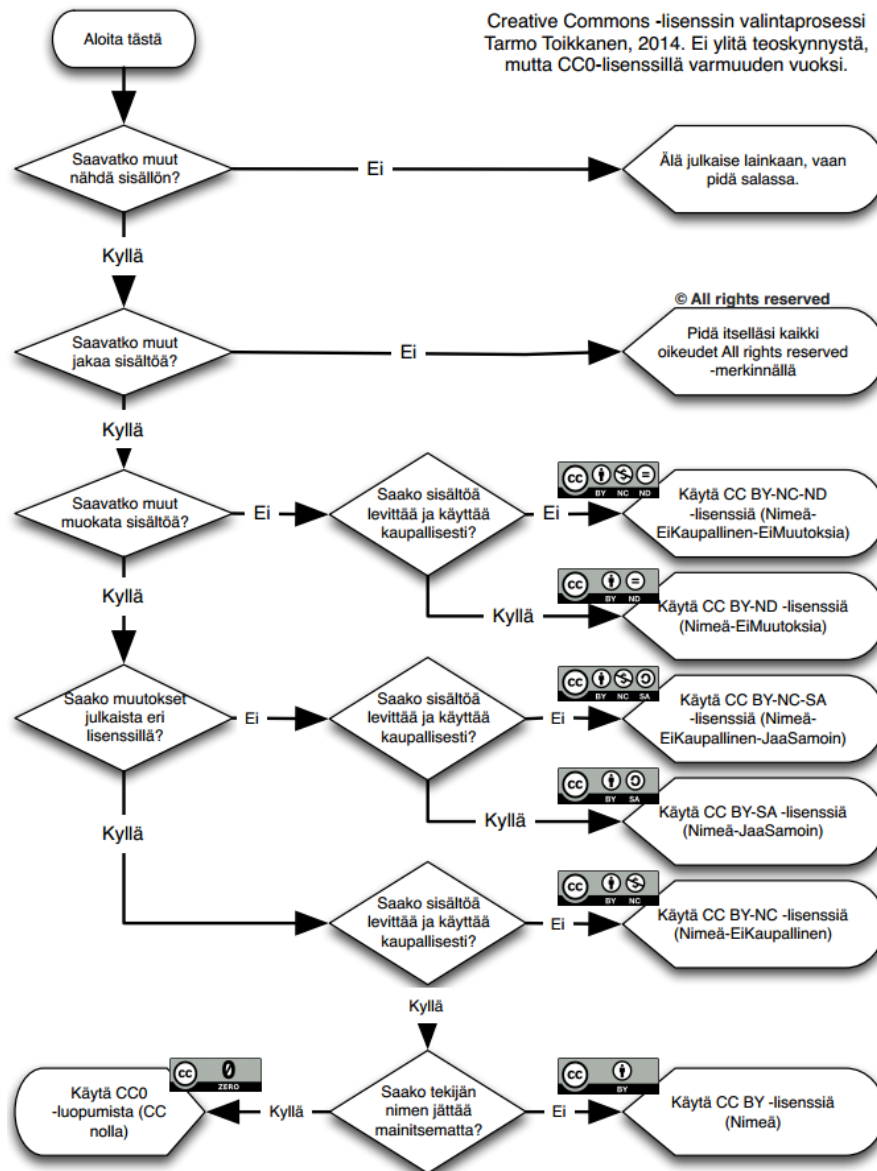
- Sallitaanko kaupallinen käyttö?
- Sallitaanko muunnelmien tekeminen?
- Jos muunnelmien tekeminen sallitaan, pitääkö ne julkaista samoilla käyttöehdoilla?

Näiden kysymysten vastausten perusteella pystytään valitsemaan oikea lisenssi, jotka ovat esitelty alla Taulukossa 1.

Taulukko 1. Creative Commonsin kuusi päälisenssiä. (Creative Commons. n.d-b.)

	Nimeä CC BY
<p>Lisenssi sallii muiden levittää, muokata ja luoda uusia teoksia aineiston pohjalta, myös kaupallisessa tarkoituksessa, kunhan aineiston alkuperäinen tekijä mainitaan. CC BY on lisensseistä kaikista joustavin.</p>	
	Nimeä-EiMuutoksia CC BY-ND
<p>Lisenssi sallii levittämisen, myös kaupallisiin tarkoituksiin kunhan aineistoa levitetään kokonaisuutena ja muuttamattomana ja alkuperäinen tekijä mainitaan.</p>	
	Nimeä-EiKaupallinen-JaaSamoin CC BY-NC-SA
<p>Lisenssi sallii muiden levittää, muokata ja luoda uusia teoksia aineiston pohjalta, kuitenkin vain epäkaupallisessa tarkoituksessa. Alkuperäinen tekijä mainittava ja uudet tuotokset on jaettava samalla lisenssillä.</p>	
	Nimeä-JaaSamoin CC BY-SA
<p>Lisenssi sallii muiden levittää, muokata ja luoda uusia teoksia aineiston pohjalta, myös kaupallisessa tarkoituksessa, kunhan aineiston alkuperäinen tekijä mainitaan ja uudet tuotokset jaetaan samalla lisenssillä.</p>	
	Nimeä-EiKaupallinen CC BY-NC
<p>Lisenssi sallii muiden levittää, muokata ja luoda uusia teoksia aineiston pohjalta, mutta vain epäkaupallisessa tarkoituksessa, kunhan aineiston alkuperäinen tekijä mainitaan. Uusienkin tuotosten tulee olla epäkaupallisia mutta johdannaisteoksia ei tarvitse julkaista samalla lisenssillä.</p>	
	Nimeä-EiKaupallinen-EiMuutoksia CC BY-NC-ND
<p>Tämä on rajoittavin kuudesta lisenssistä. Teosta saa jakaa mutta sitä ei saa muuttaa eikä hyödyntää kaupallisissa tarkoituksissa. Alkuperäinen tekijä on mainittava.</p>	

Lisäksi on olemassa CC nolla -lisenssi, joka on muuten vastaava kuin CC BY Nimeä -lisenssi, mutta edes alkuperäistä tekijää ei tarvitse mainita. Tällöin siis tekijä käytännössä luopuu kaikista oikeuksistaan. Tarkemmat kuvaukset ja oikeudelliset tekstit lisenssejä koskien löytyvät osoitteesta [www.creativecommons.org/licences](http://www.creativecommons.org/licences)



Kuva 7. Creative Commons -lisenssin valintaprosessi. (Toikkanen 2014)

### 2.3.5 Linkitetty avoin data

Nykyinen tuntemamme internet perustuu siihen, että sen alkuaikoina käyttäjät alkoivat linkittää dokumentteja ja sivustoja toisiinsa URL -osoitteiden avulla ja näin verkko alkoi kasvaa. Nykypäivänä voidaan sivustojen ja dokumenttien lisäksi linkittää myös dataa toisiinsa. Tämä muokkaa internetiä ja sen käytettävyyttä uudelleenlaiseen suuntaan. Kun tiedolla on keskinäisiä suhteita ja niitä käsitellään koneellisesti ymmärrettävässä muodossa, puhutaan semanttisesta verkosta. (Poikola ym. 2010, 72–73.)

Semanttinen verkko koostuu dokumenttien ja datan linkityksestä itseensä tai toisiin dokumentteihin. Tähän liittyy keskeisenä osana RDF-standardi. Se on malli, joka määrittelee, miten dataa kuvataan ja miten käytetyille

termeille annetaan tunniste, joka on koneellisesti ymmärrettävissä. RDF on siis tapa kuvata datan metatietoja. (Lohtander 2013, 16–18.)

Linkitetystä datasta voidaan puhua silloin, kun http-osoitteita käytetäänkin kuvaamaan asioita eikä dokumenttien välisiä suhteita. Kun tämän http-osoitteen kautta saadaan dataa tuosta kuvatusa asiasta tai data sisältää lisää http-osoitteita, jotka linkittävät asian eteenpäin, on kyse linkitetystä datasta. Kun näitä linkkejä rakennetaan datan sisälle, syntyy automaattinen ja laajeneva verkko. (Lohtander 2013, 27.)

### 2.3.6 Big Data

Avoim data voi täyttää myös Big Datan määritelmän, varsinkin jos sitä yhdistetään muihin datalähteisiin. Avoim dataa voidaan käsitellä Big Datan liitetyillä teknologioilla ihan kuin mitä tahansa muutakin dataa. Big Data on viime vuosina ollut erittäin paljon esillä myös mediassa ja siitä puhutaan kaikkialla. Usein kuulee puhuttavan molemmista yhtä aikaa samassa yhteydessä, vaikka tosiasiallisesti nämä kaksi käsitettä eivät ole millään tapaa riippuvaisia toisistaan. Jos avoin data on enemmänkin tapa toimia, big datan kohdalla painottuvat teknologiaratkaisut.

Big datasta puhutaan nimensäkin mukaisesti usein kokonsa mukaan, eli dataa on erittäin paljon. Big data määritellään usein laajemmin v-kirjaimen avulla, joista juuri määrä (volume) on ensimmäinen. Muita siihen liitettyjä v-alkuisia tunnuspiirteitä ovat monimuotoisuus (variety), nopeus (velocity) ja arvo (value). Data on nykyään yhä monimuotoisempaa, eikä sitä enää tallenneta ainoastaan merkkeinä relaatiotietokantoihin, vaan se saattaa olla myös esimerkiksi kuvaa, videota tai ääntä. Dataa saadaan käyttöön yhä nopeammalla vauhdilla, voidaan jo puhua lähes reaaliaikaisesta datan käsittelystä. Jos ennen päätöksenteon tukena olleet myyntitilastot saatiin puolivuositain tai kuukausittain, nyt niitä pystytään hyödyntämään muutamien minuuttien viiveellä. Datasta on myös onnistuttava erilaisten prosessien ja analyysien avulla louhimaan arvokasta tietoa, sillä raakadata itsessään ei ole juurikaan minkään arvoista. Kaikkea tätä yhdistää tarve uusille teknologioille ja menetelmille, sillä perinteisin keinoin sitä ei pystytä käsittelemään ja toteuttamaan. (Liikenne- ja viestintäministeriö 2014, 7.)

### 2.4 Tiedon avaaminen ja julkaiseminen

Datavarantoja voidaan ajatella raaka-aineena organisaatioille. Se kuitenkin poikkeaa kaikista muista raaka-aineista siinä, että se ei kulu tai vähene jakamalla vaan päinvastoin sen arvo saattaa vain kasvaa. Datan avaaminen voikin tehostaa sen käyttöä ja tuoda sille aivan uusia käyttötarkoituksia. Tätä kautta voidaan synnyttää uusia innovaatioita ja parantaa kustannustehokkuutta. Datavarantojen avamisessa tulee kuitenkin huomioida useita asioita, ja hyvä suunnittelu on avain onnistumiseen. (Helsinki Region Infoshare 2010a.)

### 2.4.1 Suunnittelu ja toteutus

Datavarantojen avaaminen on iso prosessi ja se vaatii paljon resursseja. Siksi toimenpidettä ei kannata yrittää suorittaa kerralla vaan pienissä osissa. On loogista, että omia tietovarantoja aletaan priorisoida oletetun tarpeen mukaan, mutta ikinä ei voi tietää varmaksi mille aineistolle todellinen kysyntä kohdistuu, sillä käyttäjien tarpeet ja sovelluskohteet voivat olla piileviä ja muodostua vasta kun data on saatavilla. (Poikola ym. 2010, 38.)

Teknisestä näkökulmasta katsottuna datavarantojen avaaminen on oman tietoarkkitehtuurin ja tietojärjestelmien kehittämistä. On mietittävä, missä formaatissa tieto on nyt ja miten se saadaan muutettua toimivampaan formaattiin, mitkä ovat käytettävissä olevat ja tarvittavat rajapinnat ja niin edelleen. Avaamisen lähestyminen teknisestä näkökulmasta voi vaikuttaa liian haastavalta työltä ja pahimmassa tapauksessa se pysäyttää koko prosessin. Onkin siis viisaampaa lähteä liikenteeseen helpomman kautta pienin askelin ja kehittää teknisiä asioita ohessa tarpeiden mukaan. (Poikola ym. 2010, 38.)

Ensimmäiseksi kannattaa kartoittaa, mitä avattavissa olevia datavarantoja jo löytyy ja julkaista listaus näistä lyhyen kuvauksen kera. Tämän perusteella käyttäjät osaavat pyytää haluamaansa dataa. Tästä tuloksena saadaan jo tietynlainen priorisointi mietittäessä jatkotoimia. Seuraavaksi aloitetaan datan avaaminen välittämättä niinkään siitä, onko se vielä järkevässä formaatissa. Kannattaa aloittaa helposti avattavista aineistoista. Myös lisensointi on mietittävä heti alussa, jotta tieto on käytettävissä. Julkaisualustoilla ei vielä alussa ole suurta merkitystä, kunhan data on kaikkien saatavilla ja siitä ilmoitetaan esimerkiksi datakatalogeissa. Tämän jälkeen prosessia voidaan lähteä iteratiivisesti toistamaan oppien samalla uutta tekniikoista ja käyttäjien tarpeista. Prosessin edetessä voidaan alkaa miettimään datan muuttamista toimivampiin formaatteihin, rajapintoja julkaisua varten, lisensoinnin tarkentamista, dokumentointia ja datan linkittämistä. Lopulta kyseeseen tulee koko tietojärjestelmäarkkitehtuurin sovittaminen avoimen datan tarpeisiin. Tällainen iteratiivinen kehittäminen on resurssitehokasta ja mielekästä ja tarjoaa käyttäjille jatkuvasti uutta lisäarvoa. Toimenpiteitä pystytään myös suunnittelemaan ja muokkaamaan esille nousseiden tarpeiden pohjalta. (Poikola ym. 2010, 38–39.)

### 2.4.2 Lainsäädäntö

On datan avaajan tehtävä varmistua siitä, että avaamisprosessilla ei rikota lainsäädäntöä ja että avattua dataa voidaan käyttää laillisesti. Tekijänoikeuslaki on yksi merkittävimmistä laeista, joka rajoittaa aineiston käyttöä. Tähän voidaankin vaikuttaa lisensointimenetelmillä, jotka esitellään aiemmin tässä opinnäytetyössä. On kuitenkin myös monia muita lakeja, jotka datan avaajan tulee ottaa huomioon, kuten tietosuojalainsäädäntö ja kilpailulainsäädäntö. Julkisuuslaki ja maksuperustelaki taas määrittelevät sitä, minkä tiedon tulee olla julkisesti saatavilla ja miten sen tiedon luovuttamisesta voidaan periä maksua. (Poikola ym. 2010, 45–48.)

Tietosuojalainsäädäntöön voidaan laskea kuuluvaksi useita eri lakeja ja säännöksiä. Yksi sen merkittävin osa on henkilötietolaki. Sen tehtävänä on suojata ihmisten yksityiselämää mutta myös edistää hyvän tietojenkäsittelytavan ja kehittämistä ja noudattamista. Henkilötietojen käyttöä käsitellään myös laissa yksityisyyden suojasta työelämässä. (Tietosuojavaltuutetun toimisto n.d, 3.) Tietoyhteiskuntakaari on 1.1.2015 alusta voimaan tullut kokoelma keskeisistä sähköistä viestintää koskevista laeista ja asetuksista. Se on luotu poistamaan päällekkäisyyksiä, ja myös sääntelyä on selvennetty. Se pitää sisällään muun muassa vanhan sähköisen viestinnän tietosuojalain ja viestintämarkkinalain. (Liikenne ja viestintäministeriö n.d.)

Hallinnon datavarantojen avoimuudesta määrää ensisijaisesti julkisuuslainsäädäntö. Julkisuuslaissa sanotaan, että jokaisella on oikeus saada tietoonsa viranomaisten julkiset asiakirjat. Luonnollisesti tietosuojalainsäädäntö rajoittaa tätä osin ja kaikki tieto ei voikaan olla julkista. Tiedot on annettava ilmaiseksi, jos ne pyydetään sähköisessä muodossa tai kun tiedon antaminen kuuluu viranomaisen tehtäviin ja velvoitteisiin. (Poikola ym. 2010, 47.) Julkisuuslaki onkin yksi merkittävä tekijä, josta löytyvät kannustimet datan avaamiselle. Kertaalleen avoimen datan periaatteiden mukaan avattu data on kaikkien saatavilla, mikä vähentää yksittäisten kyselyiden käsittelyä.

### 2.4.3 Anonymisointi

Henkilötietolainsäädäntö määrittelee hyvin tarkasti, mitä tietoja voidaan luovuttaa, kenelle, miten ja mihin tarkoitukseen. Tämä on haaste avoimelle datalle. Moni aineisto voi sisältää paljonkin hyödynnettävissä olevaa dataa, mutta koska ne sisältävät samalla yksityishenkilöiden henkilötietoja, ei niitä voida jakaa avoimesti. Henkilötietolain määräyksiensä rajoittavuus voidaan kuitenkin kiertää ja aineistosta tehdä julkaisukelpoista anonymisoimalla henkilötiedot aineistossa. Tämä tarkoittaa, että datan joukosta poistetaan kaikki tiedot, joiden perusteella ihmiset voisivat olla tunnistettavissa. (Antikainen 2014.)

Anonymisointi on prosessi, joka on tehtävä huolella ja niin turvalliseksi, että väärinkäyttöihin syyllistyvät henkilöt eivät onnistu purkamaan sitä tai ainakin sen purkaminen on tehty erittäin vaikeaksi. Viime vuosina on tullut ilmi useita vakavia henkilötietojen vuotoja, joissa rikolliset ovat saaneet käsiinsä jopa miljoonien ihmisten henkilötietoja. Tällainen tapahtuma voi olla tietoja hallinnoivalle taholle erittäin kohtalokas. (Antikainen 2014.)

Anonymisointi voidaan esittää seuraavan esimerkin avulla, joka kuvaa yhden henkilön ostotapahtuman yksinkertaistetusti. Aineistosta yleistetään henkilön tunnistamisen mahdollistavat elementit, mutta anonymisoitu aineisto on yhä käyttökelpoista tarjoten paljon informaatiota. Toisena vaihtoehtona olisi ainoastaan poistaa henkilötiedot, mutta silloin katoaisi esimerkiksi tieto asiakkaan iästä ja paikkakunnasta.



Alkuperäinen aineisto:

Maija, Meikäläinen, Nainen, 060785-145K, Esimerkkipolku 6, 13100, Hämeenlinna, 040 -123 1234, Taulutelevisio 40”, 1kpl, 699,00€, käteinen

Anonymisoitu aineisto:

Nainen, 29 vuotta, 13100, Hämeenlinna, Taulutelevisio 40”, 1kpl, 699,00€, käteinen.

## 2.5 Avoin data Suomessa

Suomessa avoimeen dataan on panostettu viimeisinä vuosina merkittävästi. Kaikilla isoimmilla kaupungeilla on jo omat avoimen datan ohjelmat, ja niiden tueksi on perustettu useita eri hankkeita ja organisaatioita. Isot kaupungit ovatkin aloittaneet aktiivisen ja järjestelmällisen tietovarantojen avaamisen. Mukaan on liittynyt myös muita julkishallinnon organisaatioita ja nykyään saatavilla onkin paljon monimuotoista avointa dataa. Lisää syntyy jatkuvasti, kun myös pienemmät kaupungit ja useat julkishallinnon organisaatiot ovat aloittaneet datavarantojen avaamisen. Seuraavaksi olisiikin toivottavaa nähdä merkittävien suomalaisten yritysten esiintulo datan avaajina.

### 2.5.1 Tietoa merkittävimmistä tiedon avaajista

Suomesta löytyy jo useita yksittäisiä ja merkittäviä tiedon avaajia. Tietyt kaupungit kuten Tampere, Oulu ja Helsinki ovat jo aktiivisia tiedon avaajia. Näitä kaupunkien tekemiä avauksia käsitellään tarkemmin datakatalogien yhteydessä. Yksittäisiä avaajia on myös useita muita. Näistä osa on esiteltynä tässä kappaleessa.

Ilmatieteen laitos on avannut merkittävän määrän säätietoa. Avointa dataa on saatavilla reaaliaikaisena, havaintojen aikasarjoina ja ennustemalleina, ja aineistot koskevat monenlaisia eri mittaustuloksia ja havaintoja kuten tutkakuvia, liikenne-, meri- ja lentosäätä, ilmastonmuutosskenaarioita sekä auringon säteilyhavaintoja. Ilmatieteen laitoksen avoin data on saatavilla sen kotisivujensa kautta. (Ilmatieteen laitos. n.d.)

Maanmittauslaitos on avannut kaikki digitaaliset maastotietoaineistonsa kaikkien saataville ilmaiseksi. Avattuna on muun muassa korkeusmalleja ja ortoilmakuvia. Karttatietoja on myös saatavilla paikkatietoikkuna.fi- ja karttapaikka.fi -palveluista. (Maanmittauslaitos. n.d.)

Liikennevirasto on kerännyt liikennettä ja liikkumista koskevaa aineistoa ja myös avannut niitä kaikkien saataville. Liikennevirastolla on olemassa suunnitelma avoimuuden kehittämiseksi, jonka mukaisesti aineistoja tullaan avaamaan myös tulevaisuudessa. Liikenneviraston avoin data jakautuu neljään eri palveluun: Digitraffic tarjoaa liikenteeseen liittyvää dataa, kuten kelikameroita ja liikenteen sujuvuustietoa. Digiroad kattaa Suomen tie- ja katuverkot ja niiden sijainti- ja ominaisuustiedot. Liikenneverkon

paikkatietoaineistot sisältävät lataus- ja katselupalveluna esimerkiksi tietoa linja-autopysäkeistä ja tasoristeyksistä. Matka.fi taas tarjoaa rajapintana saataville tietoa joukkoliikenteestä ja sen reittitiedoista. (Liikennevirasto 2015.)

Tilastokeskus on avannut StatFin tilastotietokantansa kaikkien saataville. Se sisältää merkittävän määrän suomalaista tilastotietoa lähes kaikilta toimialasektoreilta, myös liikenne- ja matkailualalta. Tietokanta-aineistojen lisäksi tilastokeskus on avannut myös opetusaineistoja ja postinumeroalueita avointa tietoa. (Tilastokeskus 2015.)

Patentti- ja rekisterihallitus on avannut rajapinnan kautta saataville kaupparekisterin kuulutustiedot sekä YTJ-tiedot. Kuulutustiedot pitävät sisällään uusien yritysten perustamiseen tai yrityksen tietojen muuttumiseen liittyvän tiedon. YTJ-tiedoista löytyy kaupparekisteriin merkityt yritysten perustiedot mutta myös tietoa muista rekistereistä kuten verotus- ja työnantajarekistereistä. (Patentti- ja rekisterihallitus. n.d.)

Visit Helsinki on Helsingin kaupungin virallinen matkailusivusto, jota ylläpitää Helsingin Markkinointi Oy. Se on hyvä esimerkki matkailuorganisaatioiden tiedon avaamisesta. Organisaatiossa on tehty linjaus matkailudatan avaamisesta ja sen pohjalta lähdetty toteuttamaan datavarantojen avaamista. Sisällöt päivittyvät tiheään, ja saatavilla on dataa esimerkiksi pääkaupunkiseudun nähtävyyksistä, kohteista ja tapahtumista. Avattuna on myös matkailutilastoja. (Visit Helsinki. n.d.)

### 2.5.2 Datakatalogeja

Suomeen on syntynyt jo useampikin merkittävä datakatalogi. Niitä on myös yhdistelty, mikä helpottaakin datan käyttäjää löytämään etsimänsä aineisto. Isoin katalogeista on avoindata.fi, joka on toteutettu osana Valtiovarainministeriön avoimen tiedon ohjelmaa. Palvelun tuottamisesta vastaa Valtion tieto- ja viestintätekniikkakeskus. Se kokoaa datasettejä myös eri katalogeista, kuten Helsinki region infosharesta (HRI) ja maanmittauslaitoksen paikkatietohakemistosta. Katalogista löytyykin yhteensä lähes 1400 eri aineistoa koskien esimerkiksi asumista, ihmisiä, koulutusta, liikennettä ja ympäristöä. Aineistoja ovat katalogissa julkaisseet muun muassa virastot ja ministeriöt, kunnat ja museot. Eniten katalogista kuitenkin löytyy HRI:n julkaisemaa pääkaupunkiseudun avointa dataa. (avoindata.fi 2015.)

Kaupungeista esimerkiksi Tampere julkaisee kotisivuillaan omaa datakatalogiaan. Se on toteutettu yhdessä Open data TRE -organisaation kanssa. Sivuston kautta voi myös jättää omia datatoiveita. Katalogista löytyy esimerkiksi kaupungin taloustietoa sekä rakennus- ja infrastruktuuritietoa. Lisäksi kaupunki toteuttaa liikennedatan avaamista yhteistyössä ITS Factoryn kanssa. Myös Oulun kaupunki on tehnyt vastaavia avauksia, ja se jakaa aineistot omilla kotisivuillaan. Myös siellä voi jättää omia datatoiveita kuten Tampereellakin. (Tampereen kaupunki 2014; Oulun kaupunki. n.d.)

Suomalaisittain mielenkiintoisena voidaan pitää myös Euroopan unionin avoimen datan portaalia. Se sisältääkin yli 8000 aineistoa ja on siis merkittävän suuri. Aineistoja on saatavilla useasta eri aiheesta. Suurin osa aineistosta koostuu EuroStatin avauksista, jotka koskevat koko Euroopan laajuista tilastotietoa. Portaalin tarkoituksena on edistää eurooppalaisen datan innovatiivista käyttöä ja mahdollistaa sen taloudellisen potentiaalın hyödyntäminen. Portaalin tavoitteena on myös edistää politiikan ja päätöksenteon avoimuutta ja läpinäkyvyyttä. Portaalia ylläpitää Euroopan unionin julkaisutoimisto. (Euroopan unionin avoimen datan portaalı 2014.)

### 2.5.3 Hankkeet ja organisaatiot

Avoimen datan ympärille on syntynyt useita hankkeita ja organisaatioita, joiden tehtävänä on edistää avoimen datan julkaisua ja käyttöä. Osa näistä on julkishallinnon hallinnoimia, mutta on myös olemassa yksityisiä tahoja ja organisaatioita ja niiden hallinnoimia projekteja. Tässä kappaleessa on esiteltynä näistä osa.

HRI eli Helsinki Region Infoshare on pääkaupunkiseudun kaupunkien yhteinen avoimen datan verkkopalvelu. Sen tehtävänä on tukea tiedon tuottajia datan avaamisessa, jakamisessa, hyödyntämisessä ja tuottamisessa. Palvelun toteuttamisesta vastaa Helsingin kaupungin tietokeskus, joka toimii yhteistyössä Forum Virium Helsingin kanssa. Palvelurahoittajina on Helsinki, Espoo, Vantaa ja Kauniainen sekä Suomen itsenäisyyden juhlarahasto. (Helsinki Region Infoshare 2010b.)

Open Data Tampere Region -projekti on luotu Tampereen seudun julkisten tietovarantojen avaamista varten. Projektilla halutaan kasvattaa avoimen datan tietämystä ja kannustaa Tampereen seudun kuntia avaamaan datavarantojaan. Lisäksi projektin tarkoituksena on luoda yhtenäisiä toimintamalleja ja aktivoida eri avoimen datan yhteisöjä. Projektista vastaavat Tampereen seudun digitaaliset sisällöt -osaamiskeskusohjelma, Hermia Group, Tampereen kaupungin tietohallintoyksikkö ja Avoin Tampere -ohjelma. (Open Data Tampere Region n.d.)

Open Knowledge Finland on voittoa tavoittelematon rekisteröity yhdistys, jonka tehtävänä on edistää avoimen datan käyttöä Suomessa. Se on osa kansainvälistä Open Knowledge -organisaatiota. Open Knowledge Finlandin omaa toimintaa on jaettu useampiin eri aihealueisiin ja niiden parissa toimiviin ryhmiin, kuten avoin oppi, avoin demokratia ja avoin tiede. Lisäksi Open Knowledge Finland on mukana monissa muiden koordinoimissa projekteissa ja sillä on myös edustus Avoimuuden neuvottelukunnassa, joka koostuu useista eri avoimen datan toimijoista. (Open Knowledge Finland n.d-c.)

6Aika on strategia, jossa on mukana Suomen kuusi suurinta kaupunkia: Helsinki, Espoo, Vantaa, Tampere, Turku ja Oulu. Strategian ylin päättävä elin on sen johtoryhmä, joka koostuu kaikkien kaupunkien edustajista. Toimintaa toteuttaa ja johtaa 6Aika-strategiatoimisto. Strategian puitteissa toiminta-alueen organisaatioille jaetaan sekä EAKR- että ESR-rahoitusta. Strategialla on kolme painopistealuetta: Avoimet innovaatioalustat, Avoin

data ja Avoin osallisuus ja asiakkuus. Strategian tavoitteena on parantaa kaupunkien tarjoamia palveluita, lisätä liiketoimintaa ja parantaa palveluiden saatavuutta, vaikuttavuutta ja tuottavuutta. (6Aika n.d.)

Apps4Finland on avoimen datan kilpailu, jossa haetaan uusia tapoja käyttää avointa dataa hyödyksi. Kilpailua on järjestetty vuodesta 2009. Nykyään sen alla on myös alueellisia kilpailuja, kuten Apps4Pirkanmaa ja Apps4Lounais-Suomi. Kilpailua järjestää joukko avoimen datan merkittäviä toimijoita kuten Forum Virium Helsinki ja Open Knowledge Finland. Lisäksi sillä on useita järjestäjä- ja haastekumppaneita ja sponsoreina merkittäviä yrityksiä, kuten IBM ja Microsoft. (Apps4Finland n.d.)

Forum Virium Helsinki on Helsingin kaupunkikonsernin osakeyhtiö. Se kehittää digitaalisia palveluita yritysten, asukkaiden, kaupungin ja muiden julkisten organisaatioiden kanssa. Sen tarkoituksena on toteuttaa Helsingin kaupungin strategiaa. Toiminta keskittyy kehityshankkeiden vetämiseen. Vaikka Forum Virium Helsinki edistää kaikkea digitaalisuutta, on sillä avoin data merkittävänä painopistealueena. (Forum Virium Helsinki n.d.)

Avoindata.net on tukisivusto, jonka kautta organisaatiot ja käyttäjät voivat esittää avoimeen dataan liittyviä kysymyksiä sekä aineistopyyntöjä. Kysymyksiin vastaamassa on aktiivinen käyttäjäkunta. Sivusto on yksityisen henkilön ylläpitämä ja sitä moderoi koko käyttäjäkunta. Palvelu on luotu Open data Tampere Region -projektissa ja sitä hyödyntää myös useat muut avoimen datan organisaatiot. Palvelun on tarkoitus toimia pilottina, ja mikäli sille osoittautuu riittävää kysyntää, se on tarkoitus liittää osaksi kansallista avoimen datan palvelua. (avoindata.net. n.d.)

AvaraS oli Tampereen teknillisen yliopiston Porin laitoksen hanke, joka toteutettiin 1.8.2013–31.12.2014. Hanke oli EAKR-rahoitteinen. Hankkeen tarkoituksena oli tuottaa avoimista tietovarannoista liiketoimintaa Satakuntaan. Hankkeen puitteissa kerättiin paljon tietoa datavarannoista ja niiden käytöstä ja julkaistiin koulutusmateriaalia. Hankkeen myötä järjestettiin myös useita tapahtumia. Materiaalit ovat saatavilla hankkeen kotisivuilta. (Tampereen teknillinen yliopisto. n.d.)

### 3 TUTKIMUSMENETELMÄT JA TUTKIMUKSEN TOTEUTUS

Tämän opinnäytetyön tutkimus on tehty käyttäen kvalitatiivista eli laadullista tutkimusotetta. Valittaessa tutkimusmenetelmää on huomioitu tutkittava ilmiö ja tutkimusongelma.

Kvalitatiivinen tutkimus perustuu sanojen ja lauseiden käyttöön ja niillä löydösten saavuttamiseen. Kvalitatiivisen tutkimuksen vastinpari on kvantitatiivinen eli määrällinen tutkimus, jonka tarkoituksena on tulkita numeroita ja tehdä johtopäätöksiä perustuen näiden tulkintojen analysointiin. Kvalitatiivinen tutkimus on syklinen ja monivivahteinen prosessi, ja sille ei ole tarkkoja määrättyjä sääntöjä, kuten kvantitatiiviselle tutkimukselle on tyypillistä. Onkin sanottu, että kvalitatiivista tutkimusta on kaikki paitsi kvantitatiivinen tutkimus. (Kananen 2008, 24.)

Kvalitatiivisessa tutkimuksessa tarkoituksena on tutkia yksittäistä ilmiötä ja luoda uusia hypoteeseja tutkimusprosessin edetessä. Tutkittavana kohteena ovat usein prosessit, joiden tutkimiseen kvantitatiiviset, määrälliset tutkimuksen menetelmät eivät sovellu, sillä prosessit ovat monimuotoisia, ja mitattavia suureita ei voida määrittää. (Kananen 2008, 25.)

#### 3.1 Tiedonkeruumenetelmät

Tutkimusta varten on kerätty tietoa kirjallisuuden ja muiden teoreettisten lähteiden lisäksi haastattelujen avulla, jotka kohdennettiin avoimen datan sekä matkailualan asiantuntijoille. Haastattelut olivat muodoltaan teema-haastatteluja ja ne suoritettiin yksilöhaastatteluina. Yksi asiantuntijoista haastateltiin verkkokonferenssisovellusta käyttäen ja loput kasvotusten. Kaikkien haastattelut tallennettiin.

Haastattelussa haastattelijä esittää kysymyksiä haastateltaville koskien näiden ajatuksia, mielipiteitä ja niin edelleen. Esitettävät kysymykset liittyvät tutkimuskysymyksiin ja niiden avulla pyritään ratkaisemaan varsinainen tutkimusongelma. Haastattelut voidaan jaotella neljään luokkaan: strukturoidut haastattelut ovat kaikkein tarkimmin ennakkoon suunniteltuja ja niissä haastateltavalle annetaan vastausvaihtoehdot valmiina. Toisena ääripäänä on lähes normaalia keskustelua vastaava avoin haastattelu. Näiden väliin jäävät puolistrukturoitu haastattelu, jossa kysymykset ovat valmiina, mutta vastausvaihtoehdot puuttuvat sekä teemahaastattelu, joka on avoin haastattelu, mutta sen etenemistä ohjaavat etukäteen päätetyt teemat eli aihealueet. Haastattelut voidaan suorittaa yksilö- tai ryhmähaastatteluina kasvotusten tai käyttäen teknisiä apuvälineitä kuten esimerkiksi puhelinta tai sähköpostia. (Kananen 2008, 73–74.)

Teemahaastattelu on yksi yleisimmistä kvalitatiivisen tutkimuksen tiedonkeruumenetelmistä. Siinä lähestytään käsiteltävää ilmiötä eri teemojen kautta. Myös teemahaastattelu voidaan pitää yksilö- tai ryhmähaastatteluina. Haastateltavien tulee liittyä tutkittavaan ilmiöön tai mikäli se ei ole mahdollista, haastatellaan henkilöitä, jotka tietävät ilmiöstä eniten. Haastateltavien määrää ei voida etukäteen määrittää, mutta heitä tulisi olla riittävästi, jotta vastaukset alkavat toistua. Teemahaastattelun tulisi alkaa yleis-

sistä asioista ja edetä kohti yksityiskohtaisempaa. Näin varmistetaan, ettei yleisempiä asioita jää käsittelemättä. (Kananen 2010, 52–55.)

### 3.2 Analysointimenetelmät

Litterointi tarkoittaa tallennetun aineiston purkamista kirjalliseen muotoon. Toinen vaihtoehto on tehdä päätelmiä ja käyttää aineistoa suoraan tallenteilta. Litterointi on tutkimuksessa yleisesti käytetty tapa, vaikka onkin melko työläs. Sanatarkkaan litterointiin voi mennä kokoneeltakin kirjoittajalta neljästä kuuteen tuntia yhtä tallennettua haastattelutuntia kohti. Tarkimmillaan litterointi tehdään erittäin yksityiskohtaisesti, huomioiden jopa puheessa ilmenevät tauot ja äännähdykset. Tällaista litterointia käytetäänkin yleensä keskusteluanalyyseissä. Tutkijan kannattaa kuitenkin miettiä, onko näin tarkalle litteroinnille tarvetta vai suorittaako litteroinnin yleisemmällä tarkkuudella. Jos haastattelun teemat ovat selkeät, voidaan litterointi tehdä ainoastaan teemaa käsittelevistä alueista. Tutkija voi tehdä myös valinnan, että ei litteroi koko dialogia vaan ainoastaan haastateltavan puheen. (Hirsjärvi & Hurme 2001, 138–142.)

Kun haastattelut oli suoritettu, tallenteet litteroitiin eli kirjoitettiin puhtaaksi. Litterointi tehtiin propositiotasolla, eli ainoastaan ydinsisältö kirjattiin ylös, kuitenkin riittävän tarkasti, jottei mitään oleellista tietoa katoasi tässä työvaiheessa. Litteroinnissa kirjalliseen muotoon purettiin ainoastaan haastateltavan puhe teema-alueista. Nämä valinnat tehtiin tutkimuksen tiukan aikataulun vuoksi mutta myös siksi, että tarkemmalla litteroinnilla ei olisi saavutettu merkittäviä lisähyötyjä. Tämän jälkeen litteroitu aineisto segmentoitiin teemoittain tekstinkäsittelyohjelmassa, eli jokainen asiakokonaisuus eroteltiin.

Sisällönanalyysin tarkoituksena on tiivistää ja selkeyttää aineistoa niin, että mitään arvokasta tietoa ei menetetä. Aineistolähtöinen laadullinen eli induktiivinen sisällönanalyysi luokitellaan kolmeen eri vaiheeseen. Ensimmäinen vaihe on aineiston pelkistäminen eli redusointi. Siinä sanantarkat alkuperäisilmaukset pelkistetään selkeiksi asioiksi. Seuraavaksi aineisto ryhmitellään eli klusteroidaan, jolloin pelkistettyjä ilmauksia yhdistellään samankaltaisuuksien mukaan. Viimeinen vaihe on abstrahointi eli käsitteellistäminen. Siinä klusteroituja luokituksia yhdistellään kohti yleisempää tasoa niin kauan kuin se on mahdollista. Klusterointi onkin siis jo osa abstrahointiprosessia. (Tuomi & Sarajärvi 2002, 110–115.)

Aineistoa analysoitiin induktiivisen sisällönanalyysin näkökulmasta. Ensin litteroitu aineisto pelkistettiin eli redusointiin. Seuraavaksi aineistoa klusteroitiin teemoittain. Kun aineisto oli ryhmitelty, sitä pelkistettiin taso kerrallaan uusiin luokkiin mahdollisimman pitkälle, kunnes saavutettiin yhteinen pääluokka.

### 3.3 Haastateltujen asiantuntijoiden esitleminen

Nina Vesterinen toimii matkailun erityisasiantuntijana työ- ja elinkeinoministeriössä elinkeino- ja innovaatio-osastolla. Hänen tehtäviinsä kuuluu muun muassa Suomen matkailupolitiikan edistäminen ja Suomen edustaminen kansainvälisissä matkailuun liittyvissä toimielimissä.

Antti Poikola on tutkija Aalto-yliopistossa, tutkimuslaitos HIIT:ssä (Helsinki Institute for Information Technology). Hän on toiminut avoimen datan parissa vuodesta 2009 ja kirjoittanut myös alan julkaisuja. Hän on myös Open Knowledge Finland ry:n puheenjohtaja.

Minna Takala toimii Hämeen liitossa aluekehitysasiantuntijana. Matkailuala on yksi hänen työnsä painopistealueista. Hän on myös Linnan pyöräverstaan yrittäjä. Aiemmin hän on työskennellyt muun muassa Nokialla ja Aalto-yliopistossa.

Kaj Järvinen toimii yrittäjänä yrityksessään Evon Luonto. Lisäksi hän on partnerina Hotelengine-nimisessä yrityksessä, joka tuottaa hotelleille varausjärjestelmien moottoreita. Hänellä on pitkä historia matkailualan ja sille palveluita tuottavien yritysten parissa, ja aiemmin hän on työskennellyt muun muassa Finnairilla ja SAS:llä.

### 3.4 Tutkimuksen luotettavuus

Tutkimuksen luotettavuutta voidaan tarkastella reliabiliteetin ja validiteetin näkökulmasta. Reliabiliteetilla tarkoitetaan tutkimustulosten toistettavuutta ja validiteetilla sitä, onko tutkimuksessa tutkittu oikeita asioita. Näitä kahta määritelmää on kuitenkin kyseenalaistettu laadullisessa tutkimuksessa, sillä ne on alkujaan luotu tarkastelemaan määrällisen eli kvantitatiivisen tutkimuksen luotettavuutta. (Tuomi & Sarajärvi. 2002, 133–134.)

Tutkimuksessa haastateltujen määrä oli pieni. Tämä ei kuitenkaan heikennä tutkimusaineiston luotettavuutta, sillä kaikki haastateltavat olivat erikseen valittuja asiantuntijoita, joilla on laaja-alainen tietämys tutkimuksen aihepiiristä. Aineistoa saatiinkin kerättyä paljon, ja saatu aineisto saturoi hyvin. Tällä tarkoitetaan, että eri lähteistä kerätty aineisto toisti samoja asioita, jolloin aineistoa voidaan pitää luotettavana.

Haastatteli litteroi saadun aineiston tuoreeltaan. Näin pystyttiin litterointivaiheessa varmistamaan, että litteroitu aineisto vastasi haastateltavan tarkoittamia merkityksiä. Vaikka aineistosta litteroitiin ainoastaan haastateltavien osuudet teemoittain ja litterointia ei tehty sanantarkasti, ei sillä ole vaikutusta aineiston luotettavuuteen tämän kaltaisessa tutkimuksessa. Tarkemmalla aineiston purkamisella ei olisi saavutettu merkittävää lisäarvoa tutkimuksen kannalta.

Aineisto analysoitiin sisällönanalyysin tunnistettuja menetelmiä käyttäen. Tarkasteltaessa tutkimustuloksia voidaan tehdä perusteltu päätelmä, että samaan tutkimustulokseen päädyttäisiin myös, mikäli tutkimus toistettaisi-

siin vastaavassa muodossa. Tätä tukevat myös aiemmin eri konteksteissa tehdyt samankaltaiset tutkimukset ja niistä saadut tulokset.



## 4 TUTKIMUSTULOKSET

Matkailualan yrityksiä on Suomessa noin 27 000 kappaletta, ja iso osa niistä on luokiteltavissa mikro- tai pk-yrityksiksi, joille usein resurssien riittämättömyys on ongelma. Digitaalisuuden ja sähköisen liiketoiminnan merkityksen lisääntyminen aiheuttaa yrityksille paljon haasteita, varsinkin juuri resurssien ja osaamisen riittävyyden kannalta. Internetissä tarjolla olevia myynti- ja markkinointikanavia on yhä enemmän, ja myös sosiaalinen media täytyy huomioida yhtenä kanavana. Yrittäjät ovat täysin työllistettyjä jo ydinliiketoiminnan suorittamisen kanssa ja tämän kaltaisten tukitoimintojen haltuun ottamiselle tai niihin tutustumiselle ei riitä aikaa.

Myös osaaminen erilaisia sähköisiä järjestelmiä ja palveluita kohtaan saattaa olla riittämätöntä. Niillä on kuitenkin yhä isompi merkitys esimerkiksi myynnissä ja markkinoinnissa, joten yritykselle olisi tärkeää hallita nämä ilmiöt. Digitaalisuuden kohdalla tulisi kuitenkin huomioida, että se ei tarjoa valmiita menestykseen johtavia ratkaisuita, vaan ainoastaan mahdollisuuden säilyttää kilpailukyky ja toimintaedellytykset muuttuvassa ympäristössä. Tämän on todennut myös Järvinen (Järvinen 2013) Talouselämän Accenturen blogissa.

Matkailuyritysten tulisi huomioida liiketoiminnassaan myös toimialan muutokset. Asiakaskenttä sirpaloituu yhä enemmän ja yksilömatkailu lisääntyy perinteisen ryhmämatkailun rinnalla. Tietynlainen suunnitteleman matkailu, eli niin kutsuttu ad hoc -matkailu, lisääntyy. Siinä asiakas ei ole tehnyt tarkkoja matkasuunnitelmia vaan tekee päätöksiä käyttämistään palveluista vasta matkan aikana. Lisäksi palvelumuotoilua tulisi kehittää huomioiden entistä paremmin kohdentamisen tarpeet ja palveluiden elämyksellisyys. Kasvua tulisi myös hakea rajapinta-ajattelun avulla muilta toimialoilta, kuten esimerkiksi hyvinvointimatkailusta. Luonnollisesti nämä haastavat muutokset tarjoavat myös mahdollisuuden kasvulle, jota voidaan edistää sekä verkostomaisella toiminnalla että myös yhteistyöllä muiden alan yrittäjien kanssa. Saavutetuilla synergiaeduilla pystyttäisiin keventämään yksittäisten yrittäjien työmäärää ja vapauttamaan resursseja ydintoimintoihin. Yhteistyöllä pystyttäisiin myös paikkaamaan puutteita osaamisen tasossa. Tässä avoimella datalla on merkittävä rooli tulevaisuudessa. Jakamalla tietoa voidaan tukea ja kehittää verkoston kaikkien toimijoiden palveluita ja liiketoimintaa.

Ennen kuin avointa dataa tiedon lähteenä kannattaa alkaa aktiivisesti miettiä, tulisikin matkailuyritysten ensin kerätä ja hyödyntää kaikki hyödyllinen omista palveluista ja järjestelmistä saatava data. Tutkimuksessa selvisi, että matkailuyritykset eivät vielä toteuta tätä kokonaisuudessaan riittävän hyvin. Varsinkin pienet matkailuyritykset, joita Suomessa on paljon, eivät kerää ja hyödynnä dataa järjestelmällisesti ja riittävän laajasti. Isommat yritykset keräävät dataa enemmän mutta myöskään ne eivät hyödynnä sitä riittävän hyvin ja saavuta kaikkia sen kautta saatavissa olevia hyötyjä. Tätä ei tietenkään voi täysin yleistää, sillä on olemassa myös edistyksellisiä pieniä yrityksiä ja toisaalta isoja yrityksiä, jotka eivät käytä dataa lainkaan niin hyvin kuin saattaisi olettaa. Yrityksillä saattaa myös olla paljonkin tietoa, mutta se on hajallaan ja formaateissa, joissa sen hyö-

dynnettävyys on huonoa. Joissain tapauksissa tieto saattaa olla ainoastaan yrittäjällä henkilökohtaisesti tiedossa, mutta sitä ei ole kirjattu ja tallennettu mihinkään muotoon. Matti Saastamoisen (Stenroos 2015) mukaan hyödyllisimmät palvelut syntyvät yleensä siitä, että dataa yhdistellään eri lähteistä. Kun siis omat datavarannot on hyödynnetty, niistä voidaan alkaa hakea entistä suurempaa hyötyä yhdistelemällä niitä esimerkiksi erilaisten avoimien data-aineistojen kanssa.

Aivan kuten koko digitalisoituminen ja sähköinen liiketoiminta, myös avoin data tarjoaa matkailuyrityksille mahdollisuuksia kehittää liiketoimintaansa. Avoin data voidaan nähdä liiketoiminnan kiihdyttäjänä ja mahdollistajana. On vielä vaikea löytää suoria siihen perustuvia liiketoimintamahdollisuuksia mutta se tulisikin huomioida enemmänkin tiettyjä prosesseja ohjaavana ja kehittäväenä tietona. Avoin data on tietynlaista perusinfrastruktuuria, mikä kehittyy jatkuvasti.

Avoimen datan hyödyntäminen on kuitenkin yrityksissä vielä hyvin alkutaipaleella. Perinteisillä toimialoilla ei vielä edes välttämättä tunnisteta koko ilmiötä. Yritysten tulisi alkaa seurata avoimeen dataan liittyvää keskustelua ja alkaa lisätä ymmärrystään sitä kohtaan. Koulutuksia ja info-tilaisuuksia avoimesta datasta järjestetään, ja ne ovatkin hyvä tapa lisätä tietämystä siitä.

Avoimeen dataan liittyy myös tietynlainen kokeilukulttuuri. Avoimella aineistolla on helppo aloittaa yritysten sisällä tutustuminen aiheeseen ja katsoa, minkälaista dataa on tarjolla ja miten sitä voitaisiin hyödyntää. Sen jälkeen voidaan aloittaa kevyet kokeilut. Nämä kokeilut eivät tietenkään automaattisesti kasvata yrityksen liikevaihtoa, mutta niiden avulla saadaan kartutettua kokemusta ja voidaan löytää hyödyllisiä käyttökohteita. Julkishallinto on avannut jo paljon dataa, mutta avaamatta on silti vielä isoja määriä aineistoja, joten välttämättä juuri se yritykselle oleellinen data ei ole vielä saatavissa. Julkishallinnon organisaatiot kuitenkin toivovat, että yritykset kertoisivat, millaista dataa haluttaisiin avattavan, joten kun yrityksessä tunnistetaan datan tarve, voidaan sitä lähteä pyytämään eri organisaatioilta.

Nykypäivänä myös teknologia mahdollistaa erilaiset kokeilut, sillä nykyään myös aivan pienimmillään yrityksillä on mahdollisuus hyödyntää laitteita ja järjestelmiä, mitkä ennen olivat isojen yritysten yksinoikeus, koska ne eivät vaadi isoja investointeja. Vaikka yritykset olisivat halukkaita aloittamaan avoimen datan hyödyntämisen ja siihen tutustumisen, esteeksi saattaa nousta ajan puute. Siksi ei tulisikaan miettiä liian suuria suunnitelmia vaan aloittaa pienestä ja tehdä toimia iteratiivisesti hiljalleen edeten. Myös Government Technology -verkkojulkaisun artikkelissa (Wood 2004) kerrotaan, että on turha huolestua kaiken kerrallaan suunnittelusta. Samassa artikkelissa viitataan myös juuri siihen, että avoin data mahdollistaa kokeilujen kautta mahdollisuuksien etsimisen.

Kun tutkimuksessa lähdettiin selvittämään yritysten roolia datan avaajana, esille nousi kaksi eri näkökulmaa. Jo valmiiksi julkista tietoa, kuten yritysten kotisivuilta löytyvää, tulisi avata ilmaiseksi koneluettavassa muodossa.

Tämä mahdollistaisi asiakkaille, sovelluskehittäjille ja muille yrittäjille tiedon monimuotoisemman käytön. Tähän yritykset myös suhtautuvat positiivisesti, koska pelkoa kilpailuedun menettämisestä ei ole. Toinen näkökulma koskee liiketoiminta- ja asiakastietoja, joilla saattaa olla vaikutusta kilpailutilanteeseen. Tätä tietoa yritykset eivät halua jakaa täysin avoimesti. Osin se on perusteltua, ja kaikkea tietoa ei voikaan jakaa, jos haluaa suojata oman kilpailuasemansa. Tästä asiasta myös Say on samoilla linjoilla kirjoittamassaan artikkelissa (Say 2014). Toisaalta tähän myös liittyy paljon asenteellisia ennakkoluuloja. Ne eivät kuitenkaan saisi vaikuttaa päätökseen datan avaamisesta. Tällaista arkaluontoisempaa dataa voitaisiin kuitenkin jakaa omien verkostojen ja sidosryhmien kanssa ilman, että se päättyy kilpailijoiden käytettäväksi. Näin kehitettäisiin kaikkien verkoston yritysten palveluita ja liiketoimintaa ilman pelkoa kilpailuedun menettämisestä. Saavutettavia hyötyjä voisi olla esimerkiksi markkinoinnin ja sen kohdentamisen kehittyminen ja palvelumuotoilun parantuminen koko verkostossa, mikä näkyisi asiakkaalle parempana kokonaispalveluna.

Myös Computerworldin artikkelissa (Finnegan 2014.) viitataan Tim Berners-Leen lausuntoon, että yritysten tulee nykypäivänä jakaa tietoa partnereiden kanssa parantaakseen tehokkuuttaan eikä niiden pidä turvata vain omaan tietoon. Yleisesti myös yritykset suhtautuisivat tähän tapaan suopeammin mutta haasteena ovat kuitenkin luottamuskysymykset, jotka ovat varsin yksilöllisiä eri yrittäjien ja yritysten kohdalla. Artikkelissaan Karvanen (Karvanen 2014) esittää huomion, että vaikka yritysten ei kannatakaan antaa kilpailijoilleen arvokasta tietoa käyttöön, voitaisiin jakaa kaikille avoimesti sellainen tieto, joka ei enää ole liiketoiminnallisesti arvokasta, eli esimerkiksi vanhentunut tieto.

Yritysten datan avaamiseen liittyy myös muita haasteita. Usein yritysten tiedot ovat varsin hajallaan ja erilaisessa muodossa kussakin järjestelmässä ja yrityksessä. Tiedon laatu saattaa myös vaihdella. Edistyksellisempien yritysten tulisi alkaa yhdessä muiden organisaatioiden kanssa miettiä datan julkaisemisen standardeja aineistojen laadun ja käytettävyyden parantamiseksi. Datan avaaminen on myös aikaa vievä ja osaamista vaativa prosessi, joten samat resursseihin liittyvät haasteet, kuten aiemmin jo mainittiin, koskevat myös datan avaamista.

Kansainvälinen konsulttiyritys Deloitte on kerännyt raporttiinsa (Deloitte 2012, 12–20) mahdollisuuksia, joita avoimen datan hyödyntäminen tuo yrityksille sekä syitä, miksi yksityisten yritysten tulisi avata tai olla avaamatta dataansa. Sen mukaan yritykset voisivat hyödyntää julkishallinnon avaamaa dataa tarjoamalla asiakkailleen siihen perustuvia uusia palveluita ja käyttämällä sitä lisäämään oman datansa käytettävyyttä ja tarkkuutta yhdistelemällä datavarantoja. Deloitteen mukaan yritykset voisivat jakamalla tietoa muille yrityksille edistää kumppanien liiketoimintaa, ymmärtää asiakasta paremmin ja parantaa yhteistyötä. Kielteisinä asioina oman datan avaamiseen liittyen se näkee juuri kilpailuedun menettämisen, mutta myös sen, että yritykset eivät ymmärrä mahdollisia hyötyjä. Myös toinen kansainvälinen konsulttiyritys, Capgemini nostaa omassa raportissaan (Tinholt 2013, 7) esille avoimen datan taloudellisina hyötyinä yrityksille

uudet liiketoimintamahdollisuudet sekä paremman päätöksenteon perustuen tarkempaan tietoon.

Avoimen datan hyödyntämiselle on olemassa Suomessa täydet edellytykset. Dataa on saatavilla avoimena jo todella paljon ja lisää tulee nopeaan tahtiin. Avoimesta datasta on tullut jo normaalia toimintaa julkishallinnon organisaatioille, ja se on myös huomioitu valtakunnan politiikassa muun muassa ohjelmakirjauksilla. Suomen avoimen datan tila kestää myös kansainvälisen vertailun ja on hyvää keskiluokkaa. Toimintaa tukevia hankkeita ja organisaatioita on perustettu, ja niissä huomioidaan myös yhä enemmän yritysten tarpeet. Tällaisissa tukitoimissa tulisi kuitenkin kiinnittää erityistä huomiota toimintaan ja sen laatuun. Koulutuksissa saavutetaan parempia tuloksia, kun ne ovat enemmän valmennuksen kaltaisia eikä vain luentotyyppejä, josta siirtovaikutukset itse tekemiseen ovat heikohkoja. Kouluttajien ja ohjaajien tietotaitotaso tulisi olla myös riittävä. Koulutuksen lisäksi tulisi panostaa tietoisuuden ja ymmärryksen lisäämiseen yrityksissä. Toisaalta taas yrityksille tarpeellisinta saattaa olla saatavilla oleva apu raskaiden työvaiheiden toteuttamiseen, sillä resurssipula saattaa estää yrityksiä tekemästä näitä itse.

## 5 JOHTOPÄÄTÖKSET JA JATKOKEHITYSEHDOTUKSET

Matkailuyrityksiä on Suomessa paljon ja suurin osa niistä on hyvin pieniä. Digitaalisuuden ja sähköisen liiketoiminnan lisääntyminen asettaa näille yrityksille haasteita, mutta samalla tuo myös mahdollisuuksia. Tämä tulisi huomioida yhä merkittävämmässä määrin yrityksen liiketoiminnassa. Tämän alle asettuu myös avoin data. Usein varsinkin pienille yrityksille haasteita lisää myös pula resursseista – yrityksillä ei välttämättä ole aikaa, rahaa eikä osaamista keskittyä mihinkään ylimääräiseksi koettuun oman ydinliiketoiminnan suorittamisen ohessa.

Myös matkailuala itsessään muuttuu: ryhmämatkailun osuus vähenee yksilömatkailun lisääntyessä. Palvelumuotoiluun tulisi kiinnittää entistä enemmän huomiota lisäten palveluiden elämyksellisyyttä ja yhdistelmäpalveluita muiden toimialojen kanssa.

Nämä matkailualaa koskevat muutokset ja haasteet huomioiden yritysten tulisi hakea hyötyjä ja ratkaisuja verkostomaisella ajattelulla, yhdessä muiden yritysten kanssa. Tämän onnistumisessa datan jakamisella kumppaneiden kanssa on iso merkitys. Vaikka yritykset eivät suhtaudukaan positiivisesti liiketoimintatiedon jakamiseen täysin avoimesti, tämän kaltaiseen verkostomaiseen yhteistyöhön ja datan jakamiseen verkoston sisällä yritykset olisivat valmiita, kunhan luottamus toimijoiden välillä on kunnossa. Myös jo valmiiksi julkisen tiedon jakaminen ilmaiseksi koneluettavassa muodossa voi tuoda yritykselle hyötyjä esimerkiksi aineiston pohjalta kehitettävien sovellusten muodossa.

Datan ja siitä saatavan tiedon merkitys matkailuyrityksille on siis suuri ja se lisääntyy jatkuvasti. Matkailuyrityksillä olisi saatavilla jo omistakin palveluista ja järjestelmistä paljon hyödyllistä dataa, mutta sitä ei vielä hyödynnetä riittävästi, varsinkaan pienissä yrityksissä.

Avoim data on edullisuutensa, saatavuutensa ja käytettävyytensä vuoksi erittäin hyödyllistä aineistoa. Yritysten tulisi käyttää avointa dataa yhä enemmän oman liiketoimintansa tukena, omien datavarantojen rinnalla. Avoimen datan tila on kuitenkin lapsenkengissä perinteisten alojen kuten matkailualan yrityksissä. Jälleen yhtenä tekijänä esille voidaan nostaa sama resurssipula mutta myös ymmärryksessä ja asenteissa olisi parannettavaa. Harva yritys vielä edes tunnistaa avointa dataa ilmiönä. Avoimella datalla kuitenkin voitaisiin saavuttaa monia hyötyjä matkailuyritysten liiketoiminnalle. Avoin data ei välttämättä mahdollista uusia suoria liiketoimintamahdollisuuksia, mutta sitä voidaan käyttää tukemaan ja kiihdyttämään yritysten eri toimintoja. Yritysten tulisikin mahdollisimman pian lähteä liikkeelle vaikka pienin askelin ilman raskasta suunnittelua, tutustuen ilmiöön ja sen mahdollisuuksiin.

Suomessa on olemassa yrityksille hyvät edellytykset hyödyntää avointa dataa. Julkishallinto on avannut paljon datavarantojaan kaikkien saataville, ja uusia avauksia tulee tiheästi. Myös poliittinen ilmapiiri avoimen datan suhteen on positiivinen. Sähköisen liiketoiminnan ympärille on luotu organisaatioita ja hankkeita yritysten tueksi. Yhä useammin niissä huomioi-

daan myös avoimen datan tarpeet. Digitaalisuuden ja sähköisen liiketoiminnan mahdollisuuksien hyödyntämiselle on siis täydet edellytykset, mutta silti yritykset eivät hyödynnä sitä riittävästi. Onkin muistettava, että on yritysten oma asia, miten ne liiketoimintaansa suorittavat. Loppujen lopuksi ne ovat itse vastuussa tuloksestaan ja tekevät päätöksensä, millä tasolla haluavat toimia ja mitkä ovat niiden tavoitteet. Yrityksiä ei siis voi väkisin pakottaa hyödyntämään avointa dataa.

Avoimen datan mahdollisuuksia matkailuyrityksille voidaan kuvata muutamien konkreettisten esimerkkien avulla. Jokaisella yrityksellä on olemassa omat tarpeensa ja kehittämisen kohteensa, joten avoin data ei sellaisenaan palvele kaikkia matkailuyrityksiä, mutta antaa käsityksen siitä, miten avointa dataa voi tulevaisuudessa hyödyntää.

Ilmatieteen laitos on avannut laajasti erilaista säähän perustuvaa dataa. Säähavaintotietoja on saatavilla kymmenen minuutin havaintotiheydellä. Tätä avointa dataa pystyttäisiin matkailualalla hyödyntämään monin tavoin. Esimerkiksi golfkentät voisivat perustaa hinnoittelunsa osin siihen. Automaattinen sähköinen varausjärjestelmä voisi huomioida varaustilanteen ja ilmatieteen laitoksen sääennusteet ja perustaa varauspalvelussa tarjottavan hinnan näihin tietoihin. Kun varausaste on matala ja luvassa on sadetta, hintoja lasketaan. Toisaalta esimerkiksi museot ja muut kulttuuri-kohteet voisivat hyödyntää sää tietoja markkinoinnissaan. Asiakkaille kannattaa markkinoida tämän kaltaisia kohteita erityisesti silloin, kun ulkona on huono sää. Ilmatieteen laitokselta on saatavilla myös paljon yksityiskohtaisempaa dataa jota voitaisiin hyödyntää, kuten aaltoennusteita surffaajille tai salamahavaintoja myrskyjen seuraajille.

Uutta yritystä tai toimipistettä perustettaessa sijoittumispäätöksen tekeminen on erittäin vaikeaa mutta samalla tärkeää. Toimipisteen sijainnin valitsemisessa voisikin käyttää monenlaista avointa dataa hyödyksi. Jos tarkoituksena on perustaa päivämatkailukohde vilkkaiden tieyhteyksien varrelle, liikenneviraston liikennetiedoista voi olla apua. Tilastokeskuksen julkaisemista matkailutilastoja vertaamalla voi saada selville, onko kaupungissa tarpeeksi majoituskapasiteettia verrattuna kävijämääriin. Aiemmin mainittuja sää tietoja voi myös käyttää sijoittumispäätöksen tukena. Maanmittauslaitoksen datasta taas saa apua, jos tarkoituksena on sijoittua hyvien ulkoilumahdollisuuksien äärelle. Usein päätös perustuu enemmän tai vähemmän yrittäjän omiin tunteuksiin, ja avoimen datan avulla päätöksen tueksi saadaan tuotua enemmän faktapohjaista tietoa.

Jos ravintolat julkaisisivat avoimena datana ruokalistansa, joka sisältää tarkemmat kuvaukset, kuten hinnat, annosten sisältämät ainesosat, ravintoarvosisällöt ja kalorimäärät, voitaisiin tästä datasta koostaa matkailijoille erilaisia sovelluksia esimerkiksi mobiililaitteille. Näin ruokavaliotaan tarkasti vahtivat asiakkaat tai allergikot voisivat vieraassakin kaupungissa helposti valita omiin kriteereihin pohjautuen hänelle parhaiten soveltuvan ravintolan. Esimerkiksi moniallergisen ihmisen saattaa olla vaikea löytää sopivaa ravintolaa, jossa voi syödä turvallisesti. Valitsemalla mobiilisovelluksesta kaikki soveltumattomat ainesosat asiakas saisi suoraan listauksen ravintoloista ja niiden annoksista, jotka hänelle soveltuvat. Luonnollisesti

tämän kaltainen sovellus toisi lisäarvoa myös niille asiakkaille, joilla ei ole erityistarpeita, mutta jotka haluavat tutustua tarjontaan nopeasti ja helposti etukäteen.

Erilaiset kuljetuspalvelut ovat tärkeitä matkailijoille. Jos kaikkien kuljetusvälineiden reaaliaikaiset sijainti-, reitti- ja hintatiedot olisivat tarjolla avoimena datana, sen pohjalta voisi luoda sovelluksen, josta matkailija voisi valita haluamansa kuljetusvaihtoehdon kohteeseen. Matkailija itse määrittäisi, onko hänelle tärkeätä hinta, nopeus vai elämyksellisyys. Esimerkiksi kaupunkipyörien tai hevoskärryjen sijainnin näyttäminen kartalla tarjoaisi vaihtoehdon taksille tai linja-autolle.

Nykypäivänä suurimmalla osalla matkaajista on älypuhelin, joka sisältää GPS-ominaisuudet. Luontomatkailukohteet voisivatkin avata ympäristöstään löytyvien nähtävyyksien, kuten historiallisten kohteiden, sijainti- ja kuvailutiedot avoimesti. Tätä dataa käyttämällä matkailija voisi tutustua näihin kohteisiin tehokkaammin kulkemalla itsenäisesti luonnossa karttapalveluita hyödyntäen. Matkailija voisi myös tehdä etukäteen reittisuunnitelmia ja merkitä palveluun lisää löytämiään kiinnostavia kohteita ja arvioida jo olemassa olevia.

Seuraavaksi olisikin syytä tehdä tapaustutkimuksia pilottiyrityksissä, missä kartoitettaisiin koko avoimen datan käyttöönoton prosessi: resurssit, koulutuksen tarve, avoimena saatavilla olevan datan tarve ja hyödyntäminen, nykyiset datavarannot ja niiden avaamisen mahdollisuudet sekä saavutetut hyödyt ja kokemukset koko prosessista. Lisäksi olisi hyvä tutkia avoimeen dataan ja sähköiseen liiketoimintaan liittyvää koulutusta: Millaista koulutusta on tarjolla, sen laatu, sisältö, tavoitteet ja se, miten yritykset saataisiin osallistumaan paremmin ja mitä konkreettisia hyötyjä ne saisivat osallistumisesta.

## LÄHTEET

- 6Aika. n.d. 6Aika – Suomen suurimpien kaupunkien yhteistyöstrategia. Viitattu 29.3.2015.  
<http://6aika.fi/painopisteet/6aika-strategiaa-toteutetaan-kolmella-painopistealueella/>
- Antikainen A. 2014. Anonymisointi työkaluna big datan hyödyntämiseen. Viitattu 1.3.2015.  
<http://www.bigdata.fi/blogi/vierailijakirjoitus/anonymisointi-tyokaluna-big-datan-hyodyntamiseen>
- Apps4Finland. n.d. Viitattu 29.3.2015. <http://www.apps4finland.fi/>
- Avoindata.fi. 2015. Taustatietoa. Viitattu 15.3.2015.  
<https://www.avoindata.fi/fi/content/taustatietoa>
- Avoindata.net. n.d. Mikä avoindata.net? Viitattu 15.3.2015.  
<http://avoindata.net/dashboard/>
- Cennoapp. 2014. Mikä on API ja miksi se on SaaS ohjelmistossa niin tärkeä? Kirjoitus blogissa 14.5.2014. Viitattu 26.2.2015  
<http://blogi.cennoapp.com/2014/05/mika-on-api-ja-miksi-se-on-saas.html>
- Creative Commons. n.d-a. Creative Commonsin tehtävä. Viitattu 25.2.2015 <http://creativecommons.fi/tietoja/>
- Creative Commons. n.d-b. Tietoa lisensseistä. Viitattu 25.2.2015  
<http://creativecommons.org/licenses/>
- Deloitte. 2012. Open data – Driving growth, ingenuity and innovation. Tulostettu 7.1.2015.  
<http://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/open-data-driving-growth-ingenuity-and-innovation.pdf>
- Euroopan unionin avoimen datan portaali. 2014. Tietoa portaalista. Viitattu 15.3.2015. <https://open-data.europa.eu/fi/data>
- Finnegan, M. 2014. Businesses must exchange private data with partners to drive efficiencies, says Berner Lee. Computerworld UK. Tulostettu 13.3.2015.  
<http://www.computerworlduk.com/news/networking/3579665/businesses-must-exchange-private-data-with-partners-to-drive-efficiencies-says-berners-lee/>
- Forum Virium Helsinki. N.d. Avointa kaupunkia rakentamassa. Viitattu 29.3.2015. <https://forumvirium.fi/esittely>
- Gruenbaum, P. 2010. “Hello World!”: A Coder’s Guide to Writing API Documentation. MSDN Magazine. November 2010. Viitattu 26.2.2015  
<https://msdn.microsoft.com/en-us/magazine/gg309172.aspx>



Helsinki Region Infoshare. 2010a. Miksi avata dataa? Viitattu 1.3.2015.  
<http://www.hri.fi/fi/avaa-dataa/miksi-avata-dataa/>

Helsinki Region Infoshare. 2010b. HRI-palvelu. Viitattu 29.3.2015.  
<http://www.hri.fi/fi/hri-projekti/>

Hirsjärvi, S. & Hurme, H. 2001. Tutkimushaastattelu: Teemahaastattelun teoria ja käytäntö. Helsinki: Yliopistopaino.

Hovi, A. 2009. Tietoarkkitehtuuri. *Systeemyö* 2/2009. 12–14. Viitattu 15.2.2015.  
[http://www.sytyke.org/wordpress/wp-content/uploads/2013/06/Systeemy%C3%B6-lehti\\_2-2009.pdf](http://www.sytyke.org/wordpress/wp-content/uploads/2013/06/Systeemy%C3%B6-lehti_2-2009.pdf)

Hovi, A., Hervonen, H. & Koistinen, H. 2009. Tietovarastot ja Business Intelligence. Helsinki: WSOY.

Hämeen ammattikorkeakoulu. n.d-a. Hakijalle. Viitattu 1.2.2015.  
<http://www.hamk.fi/hakijalle/Sivut/default.aspx>

Hämeen ammattikorkeakoulu. n.d-b. Tutkimusyksiköt. Viitattu 1.2.2015.  
<http://www.hamk.fi/tyoelamalle/tutkimusyksikot/Sivut/default.aspx>

Hämeen ammattikorkeakoulu. n.d-c. Yksiköt. Viitattu 1.2.2015.  
<http://www.hamk.fi/tietoa-hamkista/yksikot/Sivut/default.aspx>

Hämeen ammattikorkeakoulu. n.d-d. Älykkäät palvelut. Viitattu 1.2.2015.  
<http://www.hamk.fi/tyoelamalle/tutkimusyksikot/alykkaat-palvelut/Sivut/default.aspx>

Ilmatieteen laitos. n.d. Avatut ja avattavat tietoaineistot. Viitattu 15.3.2015. <http://ilmatieteenlaitos.fi/avoin-data-avattavat-aineistot>

Järvinen, J. 2013. Digitaalisuus – ainako hyvä asia? Accenturen blogi. *Talouselämä*. Tulostettu 13.3.2015.  
<http://www.talouselama.fi/kumppaniblogit/accenture/digitaalisuus++ainako+hyva+asia/a2206607>

Kananen, J. 2008. Kvali – kvalitatiivisen tutkimuksen teoria ja käytänteet. *Jyväskylän ammattikorkeakoulun julkaisuja 93*. Jyväskylän ammattikorkeakoulu. Liiketalous.

Kananen, J. 2010. Opinnäytetyön kirjoittamisen käytännön opas. *Jyväskylän ammattikorkeakoulun julkaisuja 111*. Jyväskylän ammattikorkeakoulu. Liiketoiminta ja palvelut -yksikkö.

Karvanen, J. 2013. Avoin yritysdata hyödyttäisi yrityksiä ja yliopistoja. *Tietoviikko*, 15.3.2013 (5), 10. Tulostettu 15.1.2015.  
<https://jyx.jyu.fi/dspace/bitstream/handle/123456789/42049/karvanenavoi-nyritysdata.pdf?sequence=1>

Kivekäs, O. 2014. Avoin rajapinta. Kirjoitus Blogissa 16.6.2014. Viitattu 26.2.2014. <http://otsokivekas.fi/2014/06/avoin-rajapinta/>

Liikenne- ja viestintäministeriö. n.d. Tietoyhteiskuntakaari. Viitattu 28.2.2015 <http://www.lvm.fi/web/hanke/tietoyhteiskuntakaari>

Liikenne- ja viestintäministeriö. 2014. Big datan käyttö -työryhmä, pj. Taru Rastas. Big datan hyödyntäminen - julkaisuja, 20/2014. Tulostettu 22.10.2014. <http://urn.fi/URN:ISBN:978-952-243-407-4>

Liikennevirasto 2015. Liikenneviraston avoin data. Viitattu 15.3.2015. [http://portal.liikennevirasto.fi/sivu/www/f/aineistopalvelut/avoin\\_data](http://portal.liikennevirasto.fi/sivu/www/f/aineistopalvelut/avoin_data)

Lohtander, H. 2013. Avoin data ja semanttinen verkko – yhdessä kohti älykkäämpää internetiä. Pro gradu -tutkielma. Tampereen yliopisto. Informaatiotieteiden yksikkö. Tulostettu 15.1.2015 <http://urn.fi/urn:nbn:fi:uta-1-23385>

Maanmittauslaitos. n.d. Ilmaiset aineistot. Viitattu 15.3.2015. <http://www.maanmittauslaitos.fi/avoindata>

Open Data Tampere Region. n.d. Viitattu 29.3.2015. <http://www.hermiagroup.fi/opendatatre/>

Open Knowledge. n.d. About. Viitattu 1.2.2015. <https://okfn.org/about/>

Open Knowledge Finland. 2014. Avoin julkishallinnon data – mestarikurssilaiset avasivat dataa. Viitattu 16.4.2015. <http://fi.okfn.org/2014/06/18/avoin-julkishallinnon-data-mestarikurssilaiset-avasivat-dataa/>

Open Knowledge Finland. n.d-a. Avoimen tiedon määritelmä. Viitattu 1.2.2015. <http://opendefinition.org/od/1.1/fi/>

Open Knowledge Finland. n.d-b. Visio ja arvot. Viitattu 1.2.2015. <http://fi.okfn.org/about/visiojaarvot/>

Open Knowledge Finland. n.d-c. About Open Knowledge Finland. Viitattu 29.3.2015. <http://fi.okfn.org/about/>

Oulun kaupunki. n.d. Avoin data. Viitattu 15.3.2015. <http://www.ouka.fi/oulu/oulu-tietoa/avoin-data>

Patentti- ja rekisterihallitus. n.d. Avoin data. Viitattu 15.3.2015. <http://avoindata.prh.fi/>

Peltola, V. 2014. Tämän takia ei ole avoimen datan startuppeja. Kirjoitus taloussanomien blogissa 17.11.2014. Viitattu 1.3.2014 <http://www.taloussanomien.fi/kumppaniblogit/2014/11/17/taman-takia-ei-ole-avoimen-datan-startuppeja/201415884/322>

Poikola, A., Kola, P. & Hintikka, K. 2010. Julkinen data – johdatus tietovarantojen avaamiseen, Liikenne ja viestintäministeriö 2010. Tulostettu 19.8.2014. <http://urn.fi/URN:ISBN:978-952-243-146-2>

Say, M. 2013. Can open data work for the private sector? Tulostettu 13.3.2015  
<http://www.cio.co.uk/insight/data-management/can-open-data-work-for-private-sector/>

Steenroos, P. 2015. Nyt on aika avata dataa. Tampereen seudun vetovoima. Tulostettu 13.3.2015.  
<http://tampereenseudunvetovoima.fi/uutiset/nyt-on-aika-avata-dataa>

Tampereen kaupunki. 2014. Avoin data. Viitattu 15.3.2015.  
<http://www.tampere.fi/tampereinfo/avoindata.html>

Tampereen teknillinen yliopisto. n.d. Porin laitos. AvaraS –hanke. Viitattu 15.3.2015. <https://www.tut.fi/avaras/>

Tietosuojavaltuutetun toimisto. n.d. Ota oppaaksi henkilötietolaki. Viitattu 28.2.2015  
[http://www.tietosuoja.fi/material/attachments/tietosuojavaltuutettu/tietosuojavaltuutetuntoimisto/oppaat/6Jfq8WnQ7/Ota\\_oppaaksi\\_henkilotietolaki.pdf](http://www.tietosuoja.fi/material/attachments/tietosuojavaltuutettu/tietosuojavaltuutetuntoimisto/oppaat/6Jfq8WnQ7/Ota_oppaaksi_henkilotietolaki.pdf)

Tilastokeskus. 2015. Avoimet tietokanta-aineistot. Viitattu 15.3.2015.  
<http://www.stat.fi/org/avoindata/pxweb.html>

Tinholt, D. 2013. Capgemini Consulting. The Open Data Economy – Unlocking Economic Value by Opening Government and Public Data. Tulostettu 13.3.2015.  
<http://ebooks.capgemini-consulting.com/The-Open-Data-Economy/>

Toikkanen, T. 2014. Creative Commons –lisenssin valintaprosessi. Viitattu 25.2.2014.  
<http://creativecommons.fi/wp-content/uploads/2010/06/Creative-Commons-lisenssin-valintaprosessi.pdf>

Tuomi, J. & Sarajärvi, A. 2002. Laadullinen tutkimus ja sisällönanalyysi. Helsinki: Tammi.

Työ- ja elinkeinoministeriö. 2015. Matkailu on Suomessa kasvava viientiala ja merkittävä työllistäjä. Infograafi. Viitattu 11.4.2015.  
[https://www.tem.fi/files/35515/Matkailun\\_luvut\\_infograafi\\_2015\\_fin.pdf](https://www.tem.fi/files/35515/Matkailun_luvut_infograafi_2015_fin.pdf)

Wood, C. 2014. Can Open Data Find a Business Model? Government Technology. Tulostettu 13.3.2015.  
<http://www.govtech.com/data/Can-Open-Data-Find-Business-Model.html>

## HAASTATTELUT

Järvinen, K. 2014. Yrittäjä. Evon Luonto. Haastattelu 2.4.2015.

Poikola, A. 2015. Tutkija. Aalto-yliopisto. Haastattelu 25.3.2015.

Takala, M. 2015. Aluekehitysasiantuntija. Hämeen liitto. Haastattelu 16.3.2015

Vesterinen, N. 2015. erityisasiantuntija. Työ- ja elinkeinoministeriö. Haastattelu 2.4.2015.

## TEEMAHAASTATTELUISSA KÄYTETTY HAASTATTELURUNKO

Tätä haastattelurunkoa on käytetty soveltuvin osin kunkin haastattelun tukena huomioiden haastateltavien osaamis- ja erikoistumisalueet.

### MATKAILUYRITYSTEN LIIKETOIMINNAN KEHITTÄMINEN

- Mistä liiketoiminnan kasvua tulisi hakea ja miten toimintaa tulisi kehittää?
- Mitä erilaista tietoa tämän tueksi tarvitaan?

### MATKAILUYRITYKSET JA ICT

- Millaista tietoa nyt tuotetaan ja kerätään ja miten sitä hyödynnetään?
- Matkailuyritysten resurssit, osaaminen, ymmärrys ja halu hyödyntää nykyisiä ja uusia palveluita?
- Tarvitaanko yrityksiä varten tukitoimia kuten organisaatioita ja hankkeita? Minkä tueksi?

### AVOIN DATA: TIEDON HYÖDYNTÄMINEN JA JAKAMINEN

- Mitä toimintoja avoimen datan käytöllä voitaisiin yrityksissä tehostaa?
- Miten avointa dataa tulisi alkaa käyttää?
- Kannattaako tehdä yksin vai yhdessä?
- Onko yrityksillä valmiuksia käyttää avointa dataa?
- Hyötyisivätkö matkailuyritykset jos jakaisivat tietoa kumppanien ja sidosryhmien kesken? Miten ja millaista tietoa voitaisiin jakaa?
- Voisiko tietoa jakaa myös avoimesti kaikille? Mitä tietoa ja mitä hyötyjä sillä mahdollisesti saavutettaisiin?
- Onko asiakkaan rooli muuttunut? Tulisiko asiakkaat ottaa paremmin mukaan kehitysprosesseihin?
- Millaista tietoa yritysten tulisi avata?
- Millaisia asenteita yrityksissä on avointa dataa kohtaan?

### AVOIMEN DATAN NYKYTILA JA TULEVAISUUS SUOMESSA

- Valtion ja kuntien toimet ja tahtotila?
- Yritykset avoimen datan käyttäjinä ja avaajina?
- Tulevaisuudennäkymät? Suomen taso KV-vertailussa?