



OULUN YLIOPISTO  
UNIVERSITY of OULU

# **Edellytykset avoimen datan innovaatiotoiminnalle ja niiden toteutuminen Suomessa**

Oulun Yliopisto  
Tietojenkäsittelytieteiden laitos  
Kandidaatintyö  
Mika Saarela  
21.03.2015

## Tiivistelmä

Avoin data edistää hallinnon läpinäkyvyyttä ja demokratiaa sekä tehostaa hallinnon palveluja, mutta se mahdollistaa myös uusien innovatiivisten sovellusten ja palveluiden kehittämisen. Jotta avointa dataa voitaisiin hyödyntää innovatiivisten sovellusten pohjana, tulisi avoimen datan täyttää vaatimukset datan laadun, kiinnostavuuden sekä datan helpon saatavuuden suhteen.

Edellytykset avoimen datan innovoinnin mahdollistajana ovat teknisesti laadukkaat sekä kehittäjiä kiinnostavat ja hyödylliset data-aineistot, sekä avoimet API-rajapinnat. Itse avoimen datan lisäksi merkitystä on myös sillä, miten avointa dataa pyritään edistämään, sekä miten avointa dataa tuodaan kehittäjien ulottuville. Myös julkishallintojen aktiivinen osallistuminen avoimen datan julkaisijana ja edistäjänä on tärkeää. Suomessa tilanne avoimen datan suhteen on kohtuullisen hyvä, mutta parannusta tarvitaan avoimen datan laadun, API-rajapintoihin määrän, sekä avoimen datan portaalien toiminnan osalta.

# Sisällysluettelo

Tiivistelmä .....	2
Sisällysluettelo .....	3
1. Johdanto.....	4
1.1 Tutkielman tarkoitus ja motivaatio .....	4
1.2 Tutkimusongelmat, menetelmät ja rajaukset .....	4
1.3 Tutkielman rakenne .....	5
2. Avoin data ja sen mahdollisuudet .....	6
2.1 Avoin data .....	6
2.2 Avoin data innovaatioiden mahdollistajana .....	7
2.3 Avoimen datan laatu .....	9
2.4 Arvokkaan avoimen datan ominaisuudet.....	11
2.5 Avoimen datan API-rajapinnat .....	12
2.6 Avoimen datan saatavuus ja edistäminen .....	13
3. Avoin data Suomessa .....	16
3.1 Avoimen datan innovointimahdollisuudet.....	16
3.2 Avoimen datan saatavuus ja laatu .....	17
3.3 Avoimen datan edistäminen.....	20
4. Pohdinta ja johtopäätökset.....	22
5. Yhteenveto.....	26
Lähdeluettelo.....	28

# 1. Johdanto

Avoin data on ajankohtainen ilmiö, jossa yksityinen tai valtionhallinnon organisaatio julkaisee dataa vapaasti käytettäväksi kaikille. Avoin data edistää hallinnon läpinäkyvyyttä ja demokratiaa sekä tehostaa hallinnon palveluja. Avoin data mahdollistaa myös uusien innovatiivisten sovellusten ja palveluiden kehittämisen. (Dietrich, Gray, McNamara, Poikola, Pollock, Tait & Zijlstra, 2012.)

## 1.1 Tutkielman tarkoitus ja motivaatio

Tämän tutkimuksen tavoitteena oli selvittää, mitä avoimelta datalta vaaditaan, jotta sitä voitaisiin hyödyntää uusien innovatiivisten sovellusten ja palveluiden perustana. Tutkimuksen lähestymistapana oli selvittää, mitkä asiat vaikuttavat avoimen datan innovointimahdollisuuksiin kehittäjän näkökulmasta. Tutkimuksessa perehdyttiin avoimen datan laatuun, hyödynnettävyyteen sekä siihen, kuinka avoimen datan innovointia pyritään edistämään julkishallintojen toimesta. Lisäksi tutkimuksessa pyrittiin selvittämään, miten nämä edellytykset on huomioitu Suomessa.

Avoin data ja erityisesti avoin julkishallinnon data on valtava resurssi, jota ei ole vielä suuresti hyödynnetty (Dietrich et al., 2012). Avoimen datan potentiaali on kuitenkin varsin hyvin tunnustettu ja sitä on pyritty edistämään laajasti ympäri maailmaa. Esimerkiksi Euroopan unionin julkisen tiedon direktiivi (2003/98/EC, Public sector information (PSI) Directive) on osoitus siitä, että halua avoimen datan edistämiseen on laajasti. Direktiivin tavoite on edistää eurooppalaisten yritysten kilpailukykyä määrittelemällä yhteiset käytännöt avoimen datan julkaisuun ja kustannuksiin. Direktiivissä perustellaan datan avausta myös mahdollisuutena luoda uusia työpaikkoja Euroopan unionin jäsenmaihin. (EC, 2003.) Vastaavat tavoitteet on myös INSPIRE direktiivillä (Infrastructure for Spatial Information in Europe, 2007/2/EY), joka yhtenäistää eurooppalaisten julkisten hallintojen paikkatietoaineistojen ja -palveluiden saatavuutta sekä velvoittaa viranomaiset saattamaan ne tietoverkon kautta yhteiskäyttöön (EC, 2007).

Vaikka avointa dataa pyritään aktiivisesti edistämään eri puolilla maailmaa, suuria menestystarinoita odotetaan yhä avoimen datan hyödyntämisessä (Jaakkola, Mäkinen & Eteläaho, 2014b). Tämän vuoksi on tärkeää selvittää, mitä edellytyksiä avoimelta datalta vaaditaan, jotta se mahdollistaisi innovaatiotoiminnan. Kirjallisuutta, joka keskittyisi selvittämään edellytyksiä avoimeen dataan perustuvaan innovointiin, ei löydy merkittävässä määrin, vaan tutkimukset keskittyvät avoimen datan käyttöön, sen mahdollistamiin liiketoimintamalleihin tai avoimen datan markkinointiin kehittäjille.

## 1.2 Tutkimusongelmat, menetelmät ja rajaukset

Tutkimuksen tarkoituksena oli selvittää edellytyksiä avoimen datan innovaatiotoiminnalle sekä näiden edellytyksien toteutuminen Suomessa. Tutkimuksessa pyrittiin selvittämään avoimen datan hyödyntäjän näkökulmasta ne asiat, jotka mahdollistavat avoimen datan tehokkaan hyödyntämisen ja siten myös innovatiivisten sovellusten syntymisen.

Tutkimuskysymykset tälle tutkimukselle olivat:

- Mitä avoimelta dataalta vaaditaan, jotta sitä voitaisiin hyödyntää innovatiivisissa sovelluksissa?
- Miten avoimen datan innovointia pyritään edistämään?
- Minkälaiset edellytykset Suomessa on avoimeen dataan pohjautuville innovaatioille?

Tutkimusmenetelmänä käytettiin kirjallisuustutkimusta perehtyen aikaisempaan tutkimukseen eikä empiiristä tutkimusta tässä tutkimuksessa tehty. Tutkimuksessa keskityttiin tarkastelemaan avoimen datan innovointimahdollisuuksia avoimen datan laadun, sekä sen saatavuuden kannalta, sekä kuinka avoimeen dataan liittyvää innovointitoimintaa pyritään edistämään. Avoimen datan liiketoimintamalleihin eikä myöskään avoimen datan arvoketjuihin ei tämä tutkimus ota kantaa.

### 1.3 Tutkielman rakenne

Tutkielman luvussa 2 selvitetään lähtökohdat avoimen datan laadun arviointiin tutustumalla ensin avoimeen dataan ja sen vaatimukseen sekä luodaan katsaus avoimeen dataan innovoinnin edistäjänä. Luvussa 3 luodaan katsaus siihen, kuinka avointa dataa ja sen innovointimahdollisuuksia edistetään Suomessa. Luvussa 4 tutkimuksen tuloksia tarkastellaan ja muodostetaan johtopäätökset. Luku 5 on tutkielman yhteenveto.

## 2. Avoin data ja sen mahdollisuudet

Avoin data on dataa, jota kuka tahansa voi käyttää, uudelleen käyttää ja uudelleen jakaa vapaasti. Julkishallintojen avaama avoin data edesauttaa muun muassa hallinnon läpinäkyvyyttä ja demokratiaa sekä tehostaa hallinnon palveluita, mutta se mahdollistaa myös uusien innovatiivisten sovellusten ja palveluiden tuottamisen. (Dietrich et al., 2012.)

Seuraavissa aliluvuissa perehdytään tarkemmin siihen mitä avoimelta dataalta vaaditaan laadullisesti, jotta se olisi tehokkaasti hyödynnettävissä, sekä minkälainen avoin data koetaan arvokkaaksi innovatiivisten sovellusten kannalta. Lopuksi esitellään sitä, miten avoimen datan innovointitoimintaa pyritään edistämään.

### 2.1 Avoin data

Avoin data mahdollistaa uusien innovatiivisten sovellusten sekä palvelujen kehittämisen, jotka voivat olla esimerkiksi mobiilisovelluksia tai www pohjaisia palveluja (Chan, 2013; Ubaldi, 2013). Jotta avoin data olisi kuitenkin tehokkaasti hyödynnettävissä, tulisi sen täyttää avoimelle datalle asetetut kriteerit (Dietrich et al., 2012). Avoimen datan kriteerit ovat (Dietrich et al., 2012, s. 6):

- **Saatavuus:** Datan tulee olla saatavina kokonaisena ja kustannuksiltaan kohtuullisesti hinnoiteltuna, mieluiten ilmaisena. Datan tulisi olla myös helposti saatavilla, mieluiten internetin välityksellä, sekä yleisessä dataformaatissa, jotta se olisi helpommin käytettävissä.
- **Uudelleenkäytettävissä ja jaettavissa:** Datan tulee olla saatavilla sellaisilla käyttöehdoilla, että se sallii datan uudelleenkäytön, sen yhdistämisen muihin datoihin sekä vapaan uudelleenjaettavuuden
- **Kaikille käytettävissä:** Dataa tulee voida käyttää ja jakaa sitä eteenpäin ilman rajoitteita. Esimerkiksi rajoitteet: ”sallittu vain ei kaupalliseen käyttöön” tai ”vain opetuskäyttöön” eivät ole sallittuja.

Yleisimmät avoimen datan lähteet ovat julkishallinnot johtuen niiden pyrkimyksestä avoimeen hallintoon. Esimerkkejä tällaisista datalähteistä ovat maantieteellinen karttadata, tilastodata ja säädata sekä erilaisten antureiden automaattisesti mittaamat datat. (Jaakkola, Mäkinen, Henno & Mäkelä, 2014a.) Toisaalta moni yksityisen sektorin organisaatio on ennakkoluulottomasti lähtenyt avaamaan omaa datansa saaden aikaan hyötyjä omalle liiketoiminnalleen sekä positiivista näkyvyyttä omalle toiminnalleen. Dataa avaamalla yritys harvoin saavuttaa suoraan taloudellista hyötyä. Hyötyjä voi saavuttaa esimerkiksi saamalla uusia yhteistyökumppaneita, uutta kiinnostusta yhtiön päätuotteisiin tai palveluihin, uudenlaisia liiketoimintamahdollisuuksia, sekä uusia asiakkaita tuotteilleen. (Immonen, Palviainen & Ovaska, 2014a.)

Avoin data voi toimia myös osana yrityksen liiketoimintaa. Yritys voi esimerkiksi julkaista osa datastaan avoimena, sekä tarjota eritasoisia lisenssimaksuja vastaan parempia hakurajapintoja tai lisädataa käyttäjilleen. (Ferro & Osella, 2013; Kiuru, Mäkelä & Huvio, 2012; Immonen et al., 2014a.) Useimmilla organisaatioilla ei ole aina

tarvittavaa tietämystä ja osaamista kaikkien kuviteltavissa olevien innovaatioiden luomiseen, tai ne eivät pysty tuottamaan niitä tarpeeksi nopeasti pysyäkseen muuttuvien trendien mukana. Monet yritykset avaavatkin siten osan datastaan julkiseksi, jotta niiden liiketoiminnan ympärille rakentuisi uusia palveluja hyödyntämään niiden asiakkaita. (Chan, 2013.)

Avoimen datan sekä suljetun datan välimaastossa voidaan nähdä eräänlainen harmaa alue, jossa hyödynnettävät datat kerätään www-sivuja parsimalla tai datat eivät ole tarkoituksella avoimesti saatavilla johtuen esimerkiksi huonosta datan suojelusta. Harmaan alueen datat sekä ilman lupaa parsitut datat eivät kuitenkaan täytä avoimen datan määrittystä, joten niitä ei voida laskea avoimeksi dataksi. (Jaakkola et al. 2014b.)

## 2.2 Avoin data innovaatioiden mahdollistajana

Koska julkishallinnot ovat datan julkaisijoista suurimpia, ovat niiden tuottamat data-aineistot tärkeimpiä innovoinnin edistämisessä. Avoimen datan määrittäminen ei kuitenkaan rajoita sen lähdettä, vaan avointa dataa voidaan tuottaa myös yksityisten yritysten toimesta. (Bonina, 2013.) Yksityisten yritysten data-aineistot ovatkin herättäneet mielenkiintoa ja ovat nähtävissä tulevaisuudessa yhtenä merkittävimpänä lähteenä uusille innovaatioille (Herzberg, 2014).

Yksityisten yritysten data-aineistot ovat usein elintärkeitä niiden liiketoiminnalle, eivätkä siksi voi olla avoimesti saatavilla. Vaikka vapaata pääsyä yritysten liiketoiminnan kannalta sensitiivisiin datoihin ei olisikaan, niin myös niiden matalasensitiivisessä datassa voisi olla mahdollisuuksia, josta avoimen datan yhteisö hyötyisi. (Borswara, 2014.) Esimerkkinä tällaiselle datalle Boswara (2014) mainitsee muun muassa kauppojen, postilaatikoiden, raha-automaattien tai matkapuhelintukiasemien sijaintitiedot. Myös joukkoliikenteen tuottamat aikataulutiedot sekä reaaliaikaiset liikenteen sijaintitiedot, joita tuotetaan sekä yksityisen että julkisen sektorin toimesta, nähdään potentiaalisina lähteinä innovatiivisille sovelluksille (Rohunen, Markkula, Heikkilä & Heikkilä, 2014).

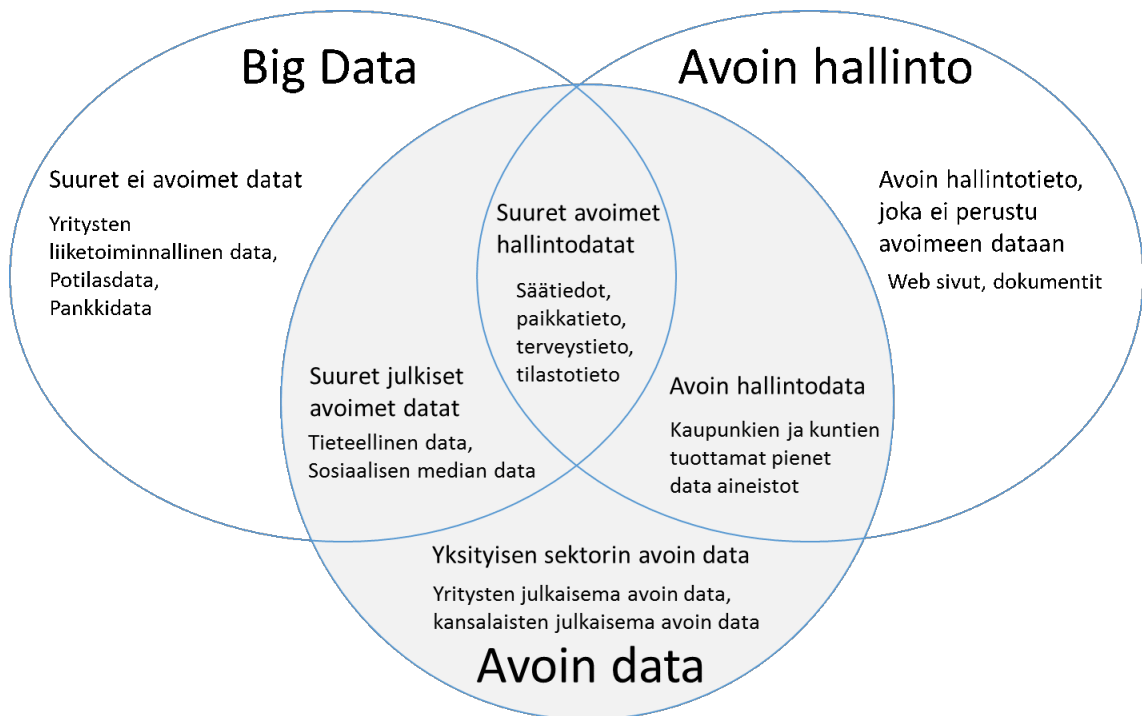
Yksityisen datan julkaisulla voidaan myös tehostaa yrityksen tuotekehitysprosessia, sillä yrityksen ulkopuoliset kehittäjät voivat kokeilla erilaisia käyttötapoja datoilta, jotka voivat siten johtaa uusiin liiketoimintamahdollisuuksiin hyödyntäen niin yritystä kuin dataa hyödyntäviä kehittäjiä. Tällaiseen kokeiluun yrityksen omat resurssit eivät välttämättä riitä ja datan avauksella saavutetaan siten säästöjä. Kuitenkaan kaikkea dataa ei tule avata, jottei yrityksen liiketoiminta vaarantuisi, mutta sopivien rajattujen datojen avulla voidaan saavuttaa liiketoiminnallisia hyötyjä. (Lindman, Kinnari & Rossi, 2014; Aitamurto & Lewis, 2011.)

Avoimelle datalle läheinen käsite on ”Big Data”, jolla tarkoitetaan suuria datamassoja, joiden prosessointiin tarvitaan valtavasti konetehoja. Big Data voi olla myös avointa dataa, niin reaaliaikaista kuin arkistodataa, jonka julkaisijoina voi olla niin julkishallinnot kuin yksityiset yrityksetkin. (Millard, 2013.) Niin avoin data kuin Big Data voivat olla myös tieteellistä dataa, jotka voivat olla julkishallintojen rahoittamaa, mutta ei näiden julkaisemaa ja ylläpitämää dataa (Gurin, 2013a). Suuria avoimia dataa julkaistaan myös erinäisten tieteellisten projektien toimesta kuten Zooniverse<sup>1</sup>, jossa

---

<sup>1</sup> <https://www.zooniverse.org>

astronominen data on kaikille käytettävissä, ja kuka tahansa voi osallistua ratkaisemaan suuria tieteellisiä ongelmia (Gurin, 2013b).



Kuva 1. Avoin datan suhteet avoimeen hallintoon, sekä Big dataan. (Muokattu Gurin, 2013a).

Kuvassa 1 havainnollistetaan avoimen datan suhdetta avoimeen hallintotietoon sekä Big dataan. Kuvan perusteella avoin data voidaan jakaa neljään eri osa-alueeseen, *yksityisen sektorin avoin data*, *avoin hallintodata* sekä niiden Big data vastineensa, *suuret julkiset yksityisen sektorin tuottamat datat* ja *suuret avoimet hallintodatat*. Näistä viimeisin eli suuret julkishallinnon tuottamat avoimet datat ovat nähtävissä potentiaalisimpana ja hyödyllisimpinä avoimeen dataan perustuvalla liiketoiminnalla, sillä julkishallinnon tuottamat kansalliset säättiedot sekä paikkatiedot ovat avoimen datan tutkimuksissa eniten käytetyt esimerkit innovatiivisten ratkaisujen taustalla. (Gurin, 2013a.)

Vaikka julkishallinnon isot datamassat nähdään yhtenä potentiaalisimpana avoimen datan esimerkkinä, ei julkishallintojen tuottaman avoimen datan tarvitse olla suurta, jotta sillä olisi innovointipotentiaalia (Gurin, 2013a). Kaupunkien ja kuntien tuottamat niin suuret kuin pienetkin data-aineistot tarjoavat hyödyllistä dataa innovatiivisiin sovelluksiin, joiden avulla voidaan ratkaista kaupunkilaisten ongelmia kuten vapaan parkkipaikan löytäminen tai julkisten kulkuvälineiden aikataulutietojen selvittäminen (Desouza & Bhagwatwar, 2012). Lokaalin avoimen datan huono saatavuus on kuitenkin sovelluskehittäjille ongelma, sillä lokaalien sovellusten yhä kasvava kysyntä vaatii monipuolista avointa lokaalia dataa (Jaakkola et al., 2014b).

Avointa kaupunkidataa käsittelevä tutkimus (Desouza & Bhagwatwar, 2012) osoittaa, että lokaalia kaupunkidataa hyödyntävillä sovelluksilla on potentiaalia suuriin menestyksiin, kun dataa on saatavilla. Osa tutkimuksen esimerkkisovelluksista ovat menestyneet lokaaleilla markkinoilla ja myöhemmin laajenneet toimimaan myös muissa kaupungeissa. Lisäksi menestyneet sovellukset ovat kannustaneet kaupungeja avaamaan lisää dataa uusien innovatiivisten sovellusten toivossa. Tutkimuksen



esimerkkisovellukset hyödynsivät joukkoliikenteen aikataulu- ja reaaliaikaista sijaintitietoa sekä julkisen palvelun dataa kuten jätteenkeräyksen aikataulutietoja. Myös data rikosraportoinneista, julkisista palveluista sekä kaikenlainen ympäristödata yhdistettynä sovelluksien itse keräämiin data-aineistoihin mahdollistivat uudenlaisia mobiilipalveluita.

Lokaalit kaupunkitason data-avaukset kannustavat kehittäjiä luomaan työkaluja ja sovelluksia, jotka parantavat kaupunkilaisten elämänlaatua. (Walravens & Ballon, 2011). Pelkän avoimen datan lisäksi myös kaupungin tarjoamia avoimia rajapintoja esimerkiksi ongelmien raportointiin voi olla saatavilla. Yhdysvalloissa ja Kanadassa käytössä olevan Open311<sup>2</sup>-rajapinta mahdollistaa käyttäjien tekemät raportoinnit mobiilisovelluksen välityksellä. Raporttien avulla pyritään parantamaan pieniä ongelmia kaupunkiympäristössä, kuten kaupungin siisteyteen tai kunnossapitoon liittyvät ongelmat. (Desouza & Bhagwatwar, 2012.)

Avoimesta datasta saadaan jalostettua uusia innovatiivisia palveluja, kun dataa yhdistellään muiden datojen kanssa. Useat kirjallisuudessa esiintyneet esimerkit avoimen datan hyödyntämisessä nojaavatkin avoimen datan lisäksi itse kerättyihin datoihin, käyttäjien keräämiin datoihin sekä maksullisiin suljettuihin datoihin. (Immonen, Palviainen & Ovaska, 2014b; Chan, 2013.) Lisäksi mobiilisovelluksissa puhelimen ominaisuudet, kuten GPS ja karttatoiminnot sekä kamera yhdistettynä avoimeen dataan antavat kehittäjille mahdollisuuden luoda uudenlaisia innovaatioita ja esittää dataa muutenkin kuin vain tekstimuotoisena informaationa (Desouza & Bhagwatwar, 2012). Tulevaisuudessa yhä useampi avoimen datan lähde perustuu automaattisten laitteiden tuottamiin mittaustuloksiin, jotka ovat myös nähtävissä tärkeinä potentiaalisina lähteinä innovatiivisille sovelluksille (Jaakkola et al., 2014a).

Avoimen datan maailmanlaajuisen tutkimusprojektin raportti Open Data Barometer pyrkii mukana olevien maiden osalta pisteyttämään, kuinka ne ovat onnistuneet omaksumaan avoimen datan hyödyt ja kuinka hyvin dataa on alettu julkaisemaan sekä millaiset vaikutukset toimilla ovat olleet. Raportti listaa innovoinnin kannalta hyödyllisimmäksi avoimen datan kategorioiksi *karttadatat*, *joukkoliikenteen aikataulutiedot*, *rikostilastot* sekä *kansainväliset pörssidatat*. Sosiaalipoliittisesti tärkeitä dataa ovat *terveyssektorin suoriutumien*, *koulutus suoriutuminen*, *ympäristötilastoinnit*, *väestötiedot* ja *maanomistustiedot*. Hallinnon tilivelvollisuuden kannalta tärkeitä dataa ovat *lainsäädäntödatat*, *vaalitulokset*, *hallinnon yksityiskohtaiset budjettidatat*, *hallinnon menot* sekä *yrittäjärekisteridatat*. (Davies, 2013.)

### 2.3 Avoimen datan laatu

Avoimen datan liiketoimintamalleissa suurimpia ongelmia on datan huono laatu, joka voi vaihdella samankin julkaisijan eri datasettien välillä. Jos avoimen datan laatuun ei kiinnitetä huomiota ja laatua pyritään parantamaan, jää avoimen datan tehokas hyödyntäminen saavuttamatta. (Gurin, 2014.) Avoimen datan huono laatu onkin koettu hyödyntämisen esteeksi ja yhteen sopimattomat dataformaattit haasteeksi avoimen datan hyödyntämiselle. Sovelluskehittäjät korostavat avoimen datan formaateissa standardeja formaatteja ja lisäksi data ja mahdollinen API rajapinta tulisi olla selkeästi dokumentoitu, jotta avoin data olisi hyödynnettävissä. (Immonen et al., 2014a; Immonen et al., 2014b.)

---

<sup>2</sup> <http://wiki.open311.org/>

Yleisesti käytössä olevassa avoimen datan viiden tähden luokittelussa pyritään pisteyttämään avointa avoin dataa sen laadun perusteella (5stardata, 2012; The Five Stars of Open Data, 2010). Luokittelussa avointa dataa voidaan luokitella 1 – 5 -asteikolla siten, että arvosana 1 tarkoittaa, että data on avoimesti saatavilla, mutta sen muoto on sellaista, ettei se ole koneluettavissa ja siten helposti hyödynnettävissä. Toisen ääripään arvosanan 5 data on sellaista, että se on helposti koneluettavissa sekä avoimessa linkitetystä formaatissa, joka sisältää useita linkkejä toisiin datoihin. Taulukossa 1 arvosanat käydään tarkemmin läpi.

**Taulukko 1.** Avoimen datan viiden tähden luokittelu (5stardata, 2012; The Five Stars of Open Data, 2010).

Pisteytys	Selitys
<b>1 tähti</b>	Data on saatavilla verkosta, mutta on luettavissa vain ihmisille eikä tulkittavissa koneellisesti, koska ovat suljetuissa formaateissa.
<b>2 tähteä</b>	Data on saatavilla verkosta ja on rakenteista koneluettavaa formaattia, joita voidaan prosessoida ja viedä uusiin formaatteihin helposti, mutta ovat suljetuissa formaateissa kuten Microsoft Word tai Excel.
<b>3 tähteä</b>	Datan käyttö ei vaadi erityistä ohjelmistoa, esimerkiksi CSV tiedoston (comma-separated value) käyttö Excel tiedoston sijasta.
<b>4 tähteä</b>	Data ei ole ainoastaan saatavilla verkosta, vaan se on verkossa linkattavissa käyttäen URI osoitetta, esimerkiksi sitä hyödyntävän mobiilisovelluksen käytettäväksi.
<b>5 tähteä</b>	Data on verkossa linkattavissa, mutta se myös viittaa sisäisesti muihin datoihin hyödyntäen näin verkostoa. Tällä tavoin linkitetyn datan arvo kasvaa eksponentiaalisesti, sillä se tulee saavutettavaksi muiden datojen kautta.

Taulukon 1 mukaisesti hyvän datan tulisi olla julkaistuna koneluettavissa formaateissa ja siten, että ne eivät vaadi jotain erityistä ohjelmistoa niiden lukemiseen. Esimerkiksi PDF tai kuvatiedostot eivät ole sopivia avoimeen dataan, sillä ne eivät ole koneluettavissa formaateissa. Myöskään tiedostoformaatit kuten Excel ja Word eivät ole hyviä vaihtoehtoja, sillä ne vaativat erityisen ohjelmiston niiden tulkitsemiseen ja saavatkin siten viiden tähden luokittelussa pistemääräksi vain kaksi pistettä. Hyvän avoimen datan tulisi olla linkattavissa sitä hyödyntävään palveluun, jolloin dataa ei tarvitse ladata erikseen sekä muuntaa käytettävämpään muotoon. Tällaiset datat saavat arvostelussa 4-5 tähteä. (5stardata, 2014; The Five Stars of Open Data, 2014.)

Tärkeintä tiedon käyttäjille on, että data on saatavilla koneluettavissa avoimissa formaateissa. Esitysmuodon avoimuus ja koneluettavuus ovat avoimen datan hyödyntämisen kannalta minimivaatimus. Pilkku- tai sarkaineroteltu data eli CVS-tiedosto, sopii parhaiten taulukkomuotoiseen dataan. Etuna on lähes jokaiselta koneelta löytyvä taulukkolaskentaohjelmisto. Monimutkaisempien tietorakenteiden esittämiseen voidaan käyttää XML-formaattia ja sen eri esitysmuotoja, kuten paikkatietojen esittämiseen tarkoitettu KML-formaatti. JSON-tiedostoformaatti on XML:ää kevyempi tapa esittää niin yksinkertaisia kuin monimutkaisia tietorakenteita ohjelmallisesti käsiteltävänä datana ja helposti siirrettävässä muodossa. Etuna JSON-formaatin käytölle on sen käsittelyn helppous nykyaikaisissa ohjelmointikielissä. (Poikola, Kola, & Hintikka, 2010.)

Data voidaan julkaista myös linkitetystä RDF-formaatissa joka mahdollistaa asioiden ja konseptien helpon linkittämisen toisiinsa. Formaatti mahdollistaa usean eri aihealueen

avoimen datan linkityksen toisiinsa kuten www sivut ovat linkitetty toisiinsa hypertextilinkkien avulla. RDF-formaatti on esimerkki viiden tähden datasta, joka mahdollistaa uusien datojen löytämisen dataa käytettäessä. (5stardata, 2012; Poikola et al., 2010.)

Vaikka data olisikin teknisesti avoimesti saatavilla, se ei aina ole julkaistu lisenssillä, joka sallisi sen vapaan käytön tai lisenssiehtoja ei ole esitetty tarpeeksi selkeästi (Jaakkola et al., 2014b). Viiden tähden luokittelussa alle asteikon menevät data-aineistot ovat datoja, jotka eivät ole saatavilla avoimella lisenssillä, vaikka itse datat olisivatkin saatavilla verkossa (The Five Stars of Open Data, 2014).

Samoin kuten avoimen lähdekoodin sovelluksilla, datalla on omistaja, jolla on oikeus määritellä säännöt ja rajoitteet sille, kuinka dataa käytetään. Yleisimmät avoimen datan julkaisuun käytetyt lisenssit ovat Creative Commons (CC) lisenssit<sup>3</sup>, *Open Data Commons (ODC)* lisenssit<sup>4</sup> sekä *Open Government* lisenssit<sup>5</sup>. Näillä lisensseillä on lisäksi useita oikeustasoja, joilla voidaan rajoittaa datan käyttöä, kieltäen esimerkiksi kaupallisen hyödyntämisen. Avoimelle datalle tulisi käyttää lisensseistä sellaisia oikeustasoja, jotka sallivat datan vapaan käytön vaatien vain alkuperäisen datalähteen mainitsemisen hyödyntävässä palvelussa. (Korn & Oppenheim, 2014; Jaakkola et al., 2014b.)

Vuoden 2013 Open data Barometer tutkimuksen mukaan avoin data ei ole vielä saavuttanut ihanteellista tilaansa sen laatuun osalta, sillä vain 6.6 % arvostelluista datoina täytti avoimen datan määritelmän ja vain 1.2 % saavutti neljä tai viisi tähteä viiden tähden arvostelussa. Monet datat, jotka olivat muuten avoimesti saatavilla, oli julkaistu rajoittavilla lisenssiehdoilla tai ilman selviä lisenssiehtoja. Yleisin dataformaatti julkaistuille datoina oli Excel ja toiseksi yleisin formaatti oli CVS- tiedosto. Vaikka datan laadussa on vielä parannettavaa, tulisi julkishallintojen silti tuoda yhä enemmän datoina avoimesti saataville. Kun data on saatu avattua, julkishallinnot voivat pyrkiä parantamaan niiden laatua ja poistamaan rajoittavia esteitä, jotka estävät datoina uudelleenkäytön. (Davies, 2013.)

## 2.4 Arvokkaan avoimen datan ominaisuudet

Julkishallintojen data on kerätty yleensä viralliseen käyttöön ilman suunnitelmaa niiden liiketoiminnallisesta uudelleenkäytöstä. Motivaatio kerätyn datoina avaamiseen tuleekin yleensä lain vaatimista syistä eikä datoina uudelleenkäyttömahdollisuuksista. Tämän vuoksi rajapinta datoina saattaa olla monimutkainen eikä datoina rakenne ole soveliaista tehokkaaseen datoina hyödyntämiseen. (Jaakkola et al., 2014b.)

Jotta data olisi arvokasta ja tehokkaasti hyödynnettävissä, siltä vaaditaan ominaisuuksia, jotka tekevät siitä arvokasta dataa. Datoina tulisi olla johdonmukaista, luotettavaa, saavutettavaa ja käytettävää. (Bonina, 2013.) Lisäksi useimmat data-aineistot ovat yleensä hyödyllisiä vain, jos ne ovat ajantasaisia ja ylläpidettyjä (Dietrich et al., 2012).

---

<sup>3</sup> <http://www.creativecommons.org>

<sup>4</sup> <http://www.opendatacommons.org>

<sup>5</sup> [www.nationalarchives.gov.uk/doc/open-government-licence/version/2/](http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/)

Myös Immonen et al. (2014b) ottavat kantaa hyvän avoimen datan ominaisuuksiin ja listaavat viisi ominaisuutta, joita hyvällä datalla on:

- **Dynaamista:** muuttuvaa dataa, joka voi olla esimerkiksi automaattisesti mitattuja mittaustuloksia
- **Luotettavaa:** jossa niin datan lähde kuin itse data on luotettavaa
- **Reaaliaikaisuus:** jatkuvasti päivittyvää dataa
- **Jatkuvuus:** takeet sille että data on saatavilla myös tulevaisuudessa
- **Integroitavuus:** dataa voidaan käyttää yhdessä omien datojen kanssa

Avoimet datalähteet voivat olla elintärkeitä niitä käyttävien sovellusten toiminnalle, joten datan luotettavuus ja jatkuvuus ovat tärkeimpiä vaatimuksia avoimen datan hyödyntämiselle. Epävarmuus datan oikeellisuudesta ja jatkuvuudesta voikin estää datan hyödyntämisen. (Bonina, 2013; Jaakkola et al., 2014b.)

Reaaliaikaisuus ja datan dynaamisuus ovat ominaisuuksia, jotka mahdollistavat hyödyllisten reaaliaikaisten mobiilisovellusten toiminnan. Esimerkkinä näistä voidaan mainita lukuisat reaaliaikaiseen dataan pohjautuvat joukkoliikennesovellukset. (Manyika, Chui, Groves, Farrel, Van Kuiken & Doshi, 2013; Desouza & Bhagwatwar, 2012.) Reaaliaikaisuuden lisäksi datan tulisi olla ajantasaista, jotta se olisi hyödyllistä. Avoimen datan hyödyllisyys onkin kyseenalaistettu, sillä useimmissa tapauksissa avoin data perustuu historialliseen dataan. (Jaakkola et al., 2014b.)

Julkishallinnon virastojen, jotka tuottavat dataa, sekä dataa käyttävien yritysten ja kansalaisten yhteinen tavoite on tehdä julkishallintojen avoimesta datasta mahdollisimman relevanttia, saavutettavaa, toimivaa ja tarkkaa. Julkishallinnot eivät pysty siihen yksin, vaan tarvitaan dialogia datan julkaisijoiden ja datan käyttäjien välille, jotta tunnistetaan tärkeimmät ja hyödyllisimmät avoimet datat ja etsitään keinoja niiden laadun parantamiseen. (Gurin, 2014, s. 64.)

## 2.5 Avoimen datan API-rajapinnat

Avointa dataa voidaan julkaista tiedostomuotoisena raakadatana sekä API-rajapintojen välityksellä (Poikola et al., 2010). Joissain tapauksissa, varsinkin suurten tietomassojen ollessa kyseessä, pelkkä tiedosto ole hyvä vaihtoehto datan julkaisuun, vaan tällöin datan julkaisu API-ohjelmointirajapinnan kautta parantaisi datan käytettävyyttä ja olisi näin helposti kytkettävissä sitä käyttävään sovellukseen. Datatiedoston julkaisu API-rajapinnan kautta antaa kehittäjälle mahdollisuuden valita osioita datasta ilman, että koko datamassaa täytyisi ladata yhdessä tiedostossa. API-rajapinnat ovat yleensä yhteydessä tietokantoihin, jotka päivittyvät reaaliaikaisesti, joten ne tarjoavat kehittäjiä kiinnostavaa ajantasaista sekä reaaliaikaista dataa. Kuitenkin raakadatan saatavuus tiedostoformaateissa on myös tärkeää, sillä se mahdollistaa riippumattomuuden alkuperäiseen datan tarjoajaan. Lisäksi kuka vain voi uudelleen jakaa ladattua datatiedostoa, eikä pelkoa datan saatavuuden menetyksestä ole. (Dietrich et al., 2012.)

API-rajapinnat voivat luoda myös hyötyjä yksityisille yrityksille. Datatiedoston avaaminen julkiseksi avointen rajapintojen välityksellä voi kiihdyttää tuotekehitysprosessia sekä luoda uusia keinoja kaupallistaa sisältöä erityisesti markkina-alueilla joita on vaikea muuten palvella. Julkaisemalla avoimia API-rajapintoja yritykset rohkaisevat ohjelmoijajayhteisöjä tekemään kokeiluja ja rakentamaan sovelluksia hyödyntäen yritysten tuottamaa sisältöä. (Aitamurto & Lewis, 2011.) Avoimet API-rajapinnat keräävät teknologian tai alustan ympärille aktiivisen kehittäjäyhteisön, joten yrityksen

datan tai palvelujen avaaminen API-rajapintojen välityksellä toimii katalyyttinä innovaatioille (Weiss & Gangadharan, 2010).

API-rajapinnat toimivat välittäjänä datan tarjoajan sekä dataa hyödyntävän sovelluksen välillä (Peres Núñez, 2014). Kun data on julkaistu vain suurena tiedostona, tulee dataa prosessoida ennen kuin sitä voidaan hyödyntää käytävässä sovelluksessa. Data tulee purkaa lähdetiedostosta sekä muuntaa uudelleenkäytettävämpään muotoon. Tällaiset vaiheet ovat raakadatan hyödyntämisessä välttämätön ja työläs vaihe varsinkin, jos datalähteen tiedostoformaatti vaihtuu ajan myötä. Toisaalta datan lataus sekä sen prosessointi voi luoda uusia näkökulmia datan hyödyntämiseen, sillä dataa voidaan analysoida tavoilla, joilla sitä ei ennen ole analysoitu ja saavuttaa siten uusia innovatiivisia ideoita datan hyödyntämisestä. (Lindman et al., 2014.)

Kun data on saatavilla API-rajapintojen välityksellä, niin yleisimpiä dataformaatteja ovat XML sekä JSON. Näistä XML (Extensible Markup Language) on merkintäkieli, jota käytetään yleisesti tiedonvälitykseen eri järjestelmien välillä. JSON (JavaScript Object Notation) -dataformaatti on toinen yleinen tiedonsiirtoformaatti, jota on helppo käyttää useissa ohjelmointiympäristöissä. API-rajapinnat ovat yleensä REST (Representational State Transfer) -rajapintoja, jossa kutsuva sovellus lähettää kutsun palvelimen URL-osoitteeseen, jolloin palvelin palauttaa pyyntöä vastaavan informaation esimerkiksi JSON- tai XML-muodossa. (Poikola et al., 2010.)

Useimmissa avoimen datan julkaisukanavissa käytetyt ohjelmistot, kuten avoimen lähdekoodin CKAN sekä kaupallinen Socrata, mahdollistavat datan julkaisun, datan etsimisen sekä visualisoinnin. Molemmat työkalut mahdollistavat myös data-aineistojen metatietojen selauksen API-rajapintojen välityksellä. Socrata tarjoaa myös työkaluja, jotka antavat avoimen datan julkaisijalle mahdollisuuden luoda automaattisesti generoituja API-rajapintoja. Tiedoston latausvaiheessa eri tiedostoformaateista, kuten Excel-tiedostoista voidaan automaattisesti luoda valmis API-rajapinta, joka mahdollistaa kyselyt dataan, ilman että koko data-aineistoa tulisi ladata. (CKAN, s.d.; SOCRATA, 2015.)

## 2.6 Avoimen datan saatavuus ja edistäminen

Yksi tärkeimmistä edellytyksistä avoimen datan hyödyntämiselle on datan löydettävyys (Jaakkola et al., 2014b). Avoimen datan käyttäjän tulee myös tietää, mitä avointa dataa on saatavilla, mistä avointa dataa voidaan ladata, ja milloin se on saatavilla, sekä kuinka usein dataa päivitetään (Immonen et al., 2014b). Datan julkaisussa olennaista on tehdä datasta mahdollisimman löydettävää ja tavoittaa kyseisestä datasta kiinnostuneet tahot. Perinteisesti data tuodaan saataville ohjelmoijien tuntemiin datakatalogeihin. Datan julkaisua voidaan myös mainostaa julkisessa mediassa, jolloin tieto avoimesta datasta saavuttaa ohjelmointiyhteisöjä ja muita datan hyödyntämisestä kiinnostuneita tahoja. (Poikola et al., 2010.)

Kun julkisen sektorin organisaatio päättää julkaista keräämäänsä dataa, tulee sen ensin siivota datasta virheelliset tiedot sekä poistaa sensitiiviset tiedot, jonka jälkeen data voidaan laittaa saataville jonkin avoimen datan portaalin välityksellä. Portaalin tulisi mahdollistaa datan etsimisen, löytämisen, datan laadun arvioinnin ja datan tarkastelun, sekä mahdollistaa tiedon saanti dataa koskevista lisensseistä. Siksi myös datan metatietojen, jotka kuvaavat dataa, tulisi olla saatavilla. (Zuiderwijk, Janssen & Davis, 2014.)

Jotta julkishallintojen avoin data voisi toimia innovaatioiden edistäjänä, tulisi julkishallintojen luoda eräänlainen sovellusalusta, jossa julkishallinnot tukevat ja pyrkivät innostamaan kolmannen osapuolen kehittäjiä tekemään sovelluksia yhteistyössä julkishallintojen kanssa. Hallinnon tulisi keskittyä infrastruktuurin tekemiseen ja siten mahdollistaa yksityisten toimijoiden ekosysteemin kehittymisen. (O' Reilly, 2011.)

Avoimen datan julkaisukanavissa datan helppo saatavuus ei yksistään riitä, vaan avoimen datan portaalien tulisi tarjota myös työkaluja, API-rajapintoja, opastusta, hyviä käytäntöjä sekä esimerkkejä datan käytöstä. (O' Reilly, 2011) Myös toteutetut palvelut ja sovellukset tulisi nostaa esille ja olla näytillä avoimen datan portaaleissa, jotta ne rohkaisisivat ja inspiroisivat muitakin kehittäjiä tekemään sovelluksia portaaleista löytyvästä datasta. Tällaisen sovellusalustan tarkoituksena on rohkaista yksityistä sektoria rakentamaan sovelluksia, joita hallinto ei edes harkitse, tai niiden resurssit eivät riitä niiden rakentamiseen. (O' Reilly, 2011; Chan, 2013.)

Niin O' Reilly (2011) kuin Chan (2013) vertaavatkin tällaista avoimen datan innovointialustaa Applen sovelluskauppaan, jossa Apple loi alustan ja puitteet, joiden avulla kuka tahansa pystyi rakentamaan innovatiivisia sovelluksia, ja jotka ovat kaikille käytettävissä sovelluskaupan välityksellä. O' Reilly (2011) toteaa myös, että lokaaleilla kaupunkitaso julkishallinnoilla on paremmat edellytykset luoda tällainen sovellusalusta kuin valtion tasolla. Esimerkkinä hän mainitsee San Franciscon<sup>6</sup> sekä Washingtonin<sup>7</sup> dataportaalit, joissa on esillä niin julkishallinnon kuin myös yksityisen sektorin kehittäjien tekemiä sovelluksia. Lisäksi hän kehottaa kaupunkia jakamaan osaamistaan, käytäntöjään ja standardejaan muiden kaupunkien kesken.

Lokaalin tason avoimen datan hankkeissa kaupunki voi ottaa aktiivisen roolin, jossa se tarjoaa kehitysalustan ja levityskanavan avoimelle datalle. Tällöin sillä ei kuitenkaan ole suoraa kontrollia tehtyihin sovelluksiin muuten kuin päättämällä, mitä dataa se avaa tai jättää avaamatta. Toisaalta kaupunki voi ottaa myös roolin, jossa se pyrkii stimuloimaan avoimen datan käyttöä, ja luomaan datansa ympärille kehittäjäyhteisön. Näin kaupunki voi kutsua kehittäjäyhteisöä rakentamaan ennalta määritettyjä sovelluksia, jotka hyödyntävät kaupunkia sekä kaupunkilaisia. (Walravens & Ballon, 2011.)

Avoimen datan innovointimahdollisuuksien edellytyksenä voidaan pitää julkishallinnon osallistuvaa roolia. Portaalien lisäksi kehittäjiä tulisi houkutella mukaan rakentamaan sovelluksia erilaisten kilpailujen sekä hackathon-tapahtumien muodossa. Myös yhteistyötä kehittäjien ja julkishallintojen välillä tulisi kehittää, jotta avoimesta datasta voidaan hyötyä molemminpuolisesti. (Millard, 2013; Chan, 2013; O'Reilly, 2011.) Dialogia datan käyttäjien ja julkaisijoiden välillä tulisi myös käydä, jotta tunnistetaan tärkeät tietoaaineistot, jotka tulisi avata hyödynnettäväksi. Myös palaute jo avatuista data-aineistoista auttaa kehittämään niiden laatua. (Poikola et al., 2010)

Avoimen datan innovaatiokilpailut ovat yleisesti käytetty keino houkutella kansalaisia kehittämään avoimesta datasta uusia sovelluksia ja kilpailujen rahapalkinnot motivoivat yhä useampia osallistumaan niihin (Desouza & Bhagwatwar, 2012). Tällaisten avoimen datan kilpailujen palkintosummat vaihtelevat pienistä summista suuriin, kuten New Yorkin kaupungin Big Apps<sup>8</sup> kilpailun 20.000 dollarin palkintosummiin (Walravens &

<sup>6</sup> San Franciscon dataportaali, <https://data.sfgov.org/>

<sup>7</sup> Washingtonin dataportaalit, [apps.dc.gov](https://apps.dc.gov/), [opendata.dc.gov](https://opendata.dc.gov/)

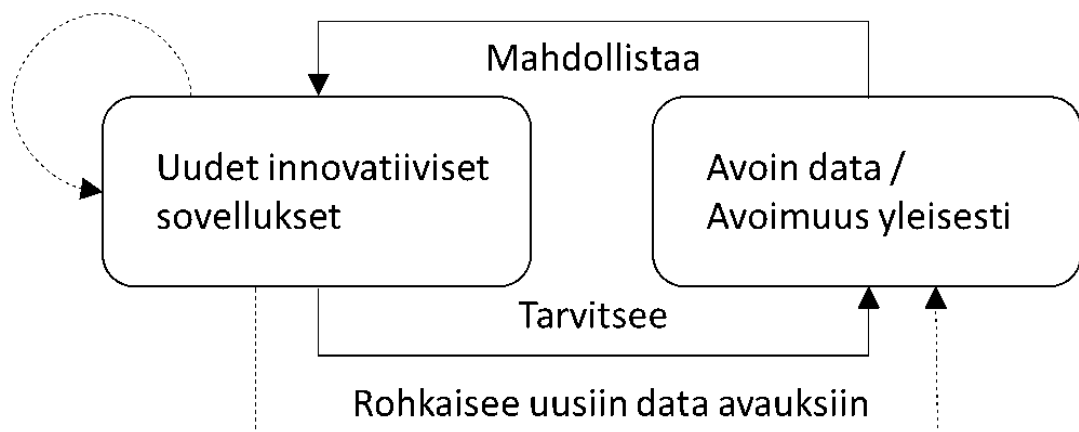
<sup>8</sup> New York Big Apps kilpailu: <http://nycbigapps.com/>

Ballon, 2011.; Desouza & Bhagwatwar, 2012). Avoimen datan innovaatiokilpailuilla on myös kauaskantoisempia hyötyjä, sillä kilpailut luovat pohjaa avoimen datan ekosysteemille ja kehittäjät tietävät, mistä avoimen datan tietolähteitä on saatavilla (Poikola et al., 2010).

Kilpailuilla on tärkeä rooli kehittäjien mukaan saamisessa. Toisaalta kilpailujen yleensä vaatimattomat palkintosummat houkuttelevat lähinnä noviiseja ammattilaisten sijasta osallistumaan tällaisiin kilpailuihin. Tämän vuoksi niissä toteutetut sovellukset ovat usein amatöörimäisiä. Myöskään kilpailun järjestäjillä ei ole kontrollia toteutettavien sovelluksien luonteeseen. Tällaiset kilpailut edistävät avoimen datan näkyvyyttä sekä laajaa yhteistyötä kehittäjien ja julkishallintojen välillä, mutta tukevat heikosti laadukkaiden ja spesifisten e-palveluiden kehittämistä. (Chan, 2013.)

Chan (2013) listaakin avoimen innovaation strategioiksi *avoimen datan innovointialustan luomisen* sekä *yhteistyöhön houkuttelemisen*. Hän lisää yhtenä houkuttelemisen keinoista formaalin yhteistyökutsun (Call-for-Collaboration), jossa ennalta määrätty tehtävä annetaan joukkoistamisen kautta vapaaehtoisille kehittäjille toteutettavaksi. Tällä tavoin julkishallinnolla on parempi kontrolli siitä, millaisia sovelluksia toteutetaan. Julkishallinto voi tarjota useita yhteistyöhankkeita, joissa käytettävä data tulee avoimen datan portaalista, ja tehtävillä sovelluksilla on ennalta määritettyjä vaatimuksia, jotka niiden tulee täyttää.

### Inspiroi



Kuva 2. Innovatiivisten sovellusten suhde avoimeen dataan.

Jaakkola et al. (2014a) kuvaavat avoimen datan suhteen innovatiivisiin sovelluksiin kaksisuuntaisena riippuvuutena, jossa avoin data mahdollistaa innovatiiviset sovellukset, jotka taas riippuvat avoimesta datasta. Nämä suhteet havainnollistetaan kuvassa 2 (Mahdollistaa ja Tarvitsee). Kuvaan on katkoviivalla lisätty kaksi uutta nuolta, joista ensimmäinen ilmaisee sitä, kuinka avoimen datan innovatiiviset sovellukset rohkaisevat uusiin data-avauksiin, kuten Desouza ja Bhagwatwarin (2012) tutkimuksen esimerkeistä kävi ilmi luvussa 2.2. Toinen nuoli (Inspiroi) kuvaa, kuinka avoimen datan innovatiiviset sovellukset voivat myös inspiroida muita kehittäjiä toteuttamaan avoimen datan sovelluksia. Toteutetut sovellukset inspiroivat muita kehittäjiä, kun sovellukset ovat julkaistu ja ovat esillä avoimen datan portaaleissa (O'Reilly, 2011; Chan, 2013; Chan, Lau & Pan, 2008). Ensimmäinen näkyvyyttä saanut sovellus tuo näkyvyyttä myös itse datalähteelle, jolloin uusia sovelluksia syntyy tämän jälkeen nopeammin (Poikola et al., 2010).

### 3. Avoin data Suomessa

Avoin data on myös Suomessa ajankohtainen ilmiö. Suomen hallitus teki vuonna 2011 periaatepäätöksen, jonka mukaan julkisen sektorin digitaalisten tietoineistojen tulee olla avoimesti saatavilla ja uudelleenkäytettävissä kaikille tasapuolisin ehdoin (Koski, Kiuru, Mäkelä & Salokannel, 2012; Kiuru et al., 2012). Avointa dataa edistetäänkin Suomessa varsin aktiivisesti, josta osoituksena on esimerkiksi Suomen avoimen hallinnon toimintasuunnitelma, jonka yhtenä tavoitteena on edistää avoimen datan saatavuutta sekä pyrkiä parantamaan jo avattujen tietoineistojen teknistä laatua. (Avoin hallinto, 2013.)

Seuraavissa aliluvuissa esitellään tarkemmin, miten avoin data näkyy Suomessa ja kuinka avointa dataa edistetään. Lisäksi perehdytään siihen, millaiset edellytykset innovoinnille on perustuen avoimen datan saatavuuteen ja laatuun.

#### 3.1 Avoimen datan innovointimahdollisuudet

Suomalaisille yrityksille tehdyn haastattelututkimuksen mukaan avoimella datalla on yritysten innovaatiotoiminnalle merkitystä, ja erityisesti pienet yritykset pystyivät hyödyntämään avattavia julkisia dataa nopeasti uusissa tuotteissaan ja palveluissaan. Kyseisen haastattelututkimuksen mukaan avoimen datan merkitys ei vielä ole tehnyt läpimurtoa, sillä sisäisten tietolähteiden merkitys nähtiin selvästi suuremmaksi innovaatioprosessin kaikissa vaiheissa. (Kiuru et al., 2012.)

Myös toisen haastattelututkimuksen mukaan avoin data on vielä uusi asia Suomessa, eikä ole vielä laajasti käytössä suomalaisissa yrityksissä, sillä vain yksi haastatelluista yrityksistä käytti avointa dataa ja sitäkin vain silloin, kun datan laatu oli riittävää hyödynnettäväksi. Kuitenkin avoimen datan potentiaali tunnustettiin haastatelluissa yrityksissä. Avoimen datan integrointi yrityksen omiin datoihin nähtiin luovan uusia mahdollisuuksia ja useimmat yritykset arvostivat nopeasti muuttuvaa dataa arvokkaampana kuin staattista muuttumatonta tai hitaasti päivittyvää dataa. (Immonen et al., 2014b.)

Suomessakin avoin data on saavuttanut menestystarinoita innovoinnin osalta. Esimerkiksi Helsingin seudun liikenne (HSL) julkaisi vuonna 2009 avoimen datan API-rajapinnan, jonka avulla kehittäjät pystyivät rakentamaan joukkoliikennesovelluksia HSL:n datan päälle. Datan avulla syntyi kymmeniä eri kolmannen osapuolen tekemiä sovelluksia, joiden toteutuksiin HSL:llä ei olisi ollut resursseja. Yrityksen ja kehittäjien välille syntyi yhteisö, jossa yritys huolehti API-rajapintojen päivityksistä kehittäjiltä saatujen palautteiden pohjalta, jotta kehittäjät pystyivät tuomaan uusia ominaisuuksia sovelluksiinsa. (Lindman et al., 2014.)

Vuonna 2012 toteutetut Maanmittauslaitoksen kartta-aineistojen avaukset ovat huomattavan laajoja ja yhteiskunnallisesti merkittäviä. Aineistojen maastotietokanta on tarkin koko maan kattava maastoa esittävä aineisto, joka sopii pohjaksi monenlaisiin maasto-sovelluksiin ja karttatuotteisiin. (Veijalainen, 2012.) Maanmittauslaitoksen avaamat maanpinnan muotoja tarkasti kuvaavat laserkeilausaineistot voivat myös houkutella tutkijoita ja yrityksiä Suomeen, sekä mahdollistavat innovatiiviset lisätyn todellisuuden (augmented reality) sovellukset (Salminen, 2013).



Avoimen datan ympärille ei ole juurikaan syntynyt liiketoimintaa, sillä kaikkiin ideoihin ei vielä löydy dataa, eivätkä demokratiasovellukset ole rahastettavissa (Salminen, 2013). Avoimen datan startup-yritysten vähäistä määrää voidaan selittää sillä, että suurin osa suomalaisesta avoimesta datasta on kehittäjien kannalta tylsää dataa. Avoin data on vielä suurimmaksi osaksi julkisen sektorin avaamia staattisia datasettejä, kuten eri virastojen budjetteja. Tulevaisuudessa on kuitenkin nähtävissä, että staattisista dataseteistä siirrytään reaaliaikaiseen virtaavaan dataan, joka on kehittäjien näkökulmasta mielenkiintoisempaa. Tällaisista datoista liikennedata on hyvä esimerkki. (Peltola, 2014.)

### 3.2 Avoimen datan saatavuus ja laatu

Open Data Barometerin<sup>9</sup> tilastojen sekä Global Open Data Indexin<sup>10</sup> mukaan Suomen tilanne on varsin hyvä karttadatojen, julkisen liikenteen avoimien datojen, vaalitulosten sekä kansallisten tilastojen suhteen. Molemmissa näistä tutkimuksista arvostellaan eri datakategorioiden osalta, miten hyvin ne vastaavat avoimen datan määritelmää huomioiden muun muassa lisensoinnit, koneluettavuuden ja ajantasaisuuden. Arvosteluissa on mukana paras esimerkki kyseisestä datasta, joten kaikkia kategorian dataa ei kummasakaan tutkimuksessa käsitellä.

Suomalaiset julkishallinnot tuottavat laadukkaita tietoaaineistoja ja ovat kohtuullisen nopeita julkaisemaan niitä (Lindman et al., 2014). Suurin osa julkaistuista aineistoista ei kuitenkaan ole koneluettavassa muodossa, joten datan ohjelmallinen ja automatisoitu hakeminen, käsittely ja yhdistely toisiin aineistoihin ovat työläitä tai käytännössä mahdottomia toteuttaa. Pyrkimystä datan koneluettavuuden parantamiseen kuitenkin on olemassa, ja datan julkaisu useammassa koneluettavassa formaatissa sisältyy avoimen datan julkaisun ohjeistuksiin. (Poikola et al., 2010.)

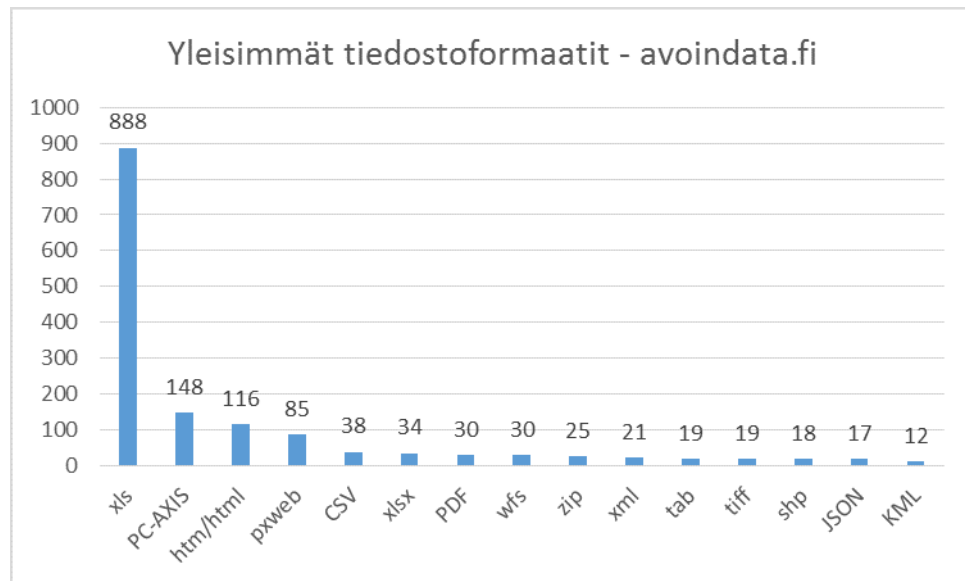
Tunnetuin suomalainen avoimen datan julkaisukanava on Avoindata.fi-portaali, joka on toteutettu osana valtiovarainministeriön avoimen tiedon ohjelmaa. Sen tavoitteena on luoda edellytykset julkishallintojen organisaatioiden avointen tietoaaineistojen avaamiseen sekä niiden käytön tehostaminen. Tietolähteitä palvelussa on 1557 kappaletta, joista suurin osa on historiallista tilastodataa. Portaalin toteutus perustuu avoimeen lähdekoodiin perustuvaan CKAN-sovellusalustaan, joka mahdollistaa datan hakuominaisuudet sekä data-aineistojen etsimisen ja metatietojen tarkastelun API-rajapintojen avulla. (avoindata.fi, 2015.)

Avoindata.fi-portaalin avoimista datalähteistä tehtyjä sovelluksia ei ole esillä palvelussa, joten esimerkkejä datalähteiden käytöstä ei ole saatavilla. Sivustolla olevien tietolähteiden hakuominaisuudet on toteutettu tekstihaun lisäksi aihealueittain, sisältötyypeittäin, tiedon tuottajittain sekä tiedostomuodon perusteella. Jokaisessa data-aineistossa on esillä viiden tähden luokittelun mukainen pisteytys aineiston laadusta sekä tieto datan lisenssistä. Pisteytys ei kuitenkaan ole ajantasaista, sillä listatuista avoimista datoista 368 kappaleella ei ole pisteytystä lainkaan. Lisäksi korkein tähtimäärä portaalin datoista on kolme, vaikka joukossa on dataa, jotka ovat linkitettävissä käyttävään sovellukseen ja olisivat näin vähintään neljän tähden arvostelun mukaisia. Jotkin data-aineistot on portaalissa arvosteltu yhden tähden

<sup>9</sup> <http://opendatabarometer.org/>

<sup>10</sup> <http://index.okfn.org/>

arvoiseksi, vaikka ne ovat saatavilla avoimella lisenssillä ja linkitetyssä dataformaattissa. Toisaalta 3 tähteä ovat saaneet myös sellaiset data-aineistot, jotka eivät ole saatavilla avoimella lisenssillä tai lisenssistä ei ole tietoa. (avoindata.fi, 2015.)



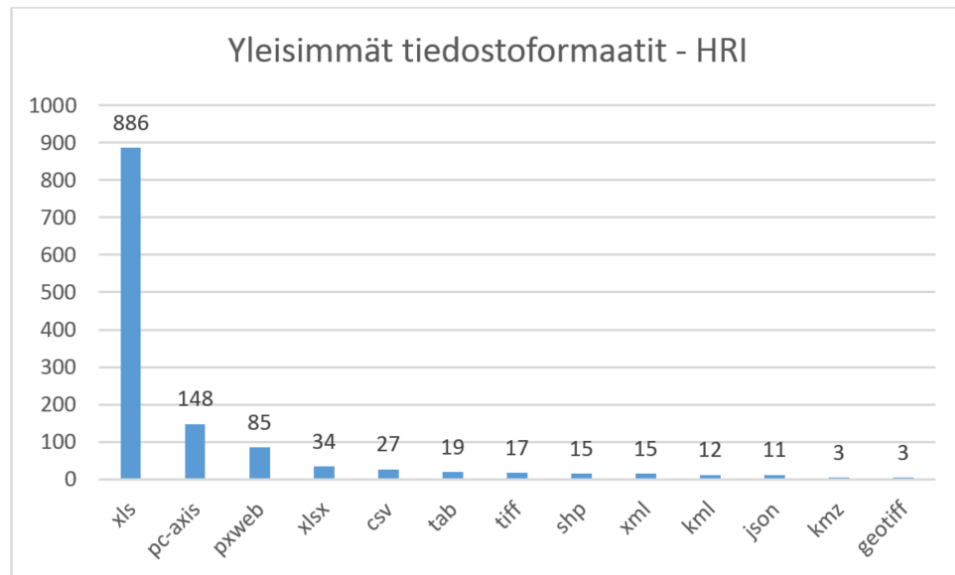
Kuva 3. Yleisimmät dataformaattit avoindata.fi dataportalissa.

Kuvan 3 mukaisesti yleisimmät tiedostotyyppit palvelussa ovat taulukkolaskentaformaattit XLS, XLSX ja CSV sekä tilastoinnissa yleisesti käytettävät formaattit (PC-AXIS ja PXWEB)<sup>11</sup>. Yleisimpien dokumenttien joukossa on myös HTM- ja HTML-dokumentit. JSON- ja XML-formaatissa saatavat datat ovat rajapinnan kautta saatavilla, joista osassa on tarjolla molemmat formaattit. Yhteenlaskettuja xml- tai JSON-rajapintoja palvelu tarjoaa 32 kappaletta. (avoindata.fi, 2015.)

Pääkaupunkiseudun lokaalin datan verkkopalvelu Helsinki Region Infoshare (HRI)-verkkopalvelu sisältää kattavasti pääkaupunkiseudun kaupunkien tuottamat tai keräämät avoimet tietoaineistot. Sivustolla olevat data-aineistot ovat osittain samat kuin avoindata.fi-portaalissa. Tietolähteitä palvelussa on 1160 kappaletta, joista suurin osa on tilastodataa samoin kuin avoindata.fi-portaalissa. Osa palvelussa olevasta datasta on tarjolla tiedostomuotoisena ladattavana datana sekä ohjelmointirajapintojen kautta käytettävänä datana. (HRI, 2015.)

Myös HRI-verkkopalvelu perustuu avoimen lähdekoodin CKAN-sovellusalustaan, joten hakuominaisuudet datalle ovat samanlaiset kuin avoindata.fi-portaalissa. HRI-palvelussa avoimen datan etsiminen on toteutettu tekstihaun lisäksi muun muassa kaupungeittain, kategorioittain sekä tiedostomuodon perusteella. Data-aineistojen tiedoissa ei ole tietoa viiden tähden pisteytyksestä. Lisensointitiedot ovat kuitenkin hyvin esillä, ja yleisin data-aineistojen lisenssi on CC-BY, joka sallii vapaan käytön, kunhan tietolähde mainitaan käyttävässä sovelluksessa. Palvelussa on myös sovellusgalleriaosio, jossa tehdyt sovellukset ovat esillä ja joiden kuvauksissa on tiedot siitä, mitä data-aineistoa niissä on käytetty. (HRI, 2015.)

<sup>11</sup> [http://www.scb.se/sv/\\_/PC-Axis/Programs/](http://www.scb.se/sv/_/PC-Axis/Programs/)



Kuva 4. Yleisimmät dataformaattit HRI dataportalissa.

Kuvan 4 mukaisesti yleisimmät tiedostomuodot HRI-palvelussa ovat pääasiassa samoja, kuin avoindata.fi-portaalissa. Suurin osa Avoindata.fi-portaalin datoista ovatkin pääkaupunkiseudun alueen dataa ja näin myös saatavilla HRI-palvelussa. (HRI, 2015; avoindata.fi, 2015.)

HRI-palvelun lisäksi paikallista kaupunkien avointa dataa on saatavilla myös useiden suomalaisten kaupunkien kotisivujen kautta, kuten Jyväskylän<sup>12</sup>, Oulun<sup>13</sup>, Tampereen<sup>14</sup>, ja Vantaan<sup>15</sup> kaupunkien avoimen datan sivuilta. Näistä Tampereen datakatalogi tarjoaa määrällisesti eniten data-aineistoja sekä monipuolisimmat vaihtoehdot tiedostomuodoista. Myös useisiin datoihin on pääsy rajapintojen välityksellä. Kaikissa edellä mainituissa paikallisissa datakatalogeissa data-aineistojen määrät ovat vielä hyvin pieniä verrattuna HRI:n dataportaliin. Myös portaalien toteutuksissa on puutteita kattavien hakutoiminnallisuuksien puuttuessa sekä vajaiden ja puuttuvien metatietojen osalta.

Avoimen datan eri portaalien lisäksi Tampereen teknillisen yliopiston Porin laitoksen vuonna 2013 aloitetun AVARAS<sup>16</sup>-projektin yhtenä tuotoksena syntynyt avoimen datan sivusto pyrkii tuomaan avoimen datan hyödyntämiseen tarvittavia resursseja kehittäjien ja yritysten tietoisuuteen. Sivusto ei itsessään listaa avoimia dataa, mutta se tarjoaa linkkejä eri avoimen datan tietolähteisiin kattavasti koko Suomesta. Sivusto pyrkii tarjoamaan myös opastusta avoimeen dataan liittyvistä teknologioista sekä työkaluista. (Jaakkola et al., 2014b.)

<sup>12</sup> <http://data.jyvaskyla.fi/>

<sup>13</sup> <http://www.ouka.fi/oulu/oulu-tietoa/avoin-data>

<sup>14</sup> <http://www.tampere.fi/tampereinfo/avoindata.html>

<sup>15</sup> [http://www.vantaa.fi/fi/tietoa\\_vantaasta/avoin\\_data](http://www.vantaa.fi/fi/tietoa_vantaasta/avoin_data)

<sup>16</sup> <https://www.tut.fi/avaras/>

Avoimen datan API-rajapintoja Suomessa on saatavilla avoimen datan portaaileista kuten avoindata.fi- sekä HRI-palveluista (HRI, 2015; avoindata.fi, 2015). Myös API Suomi-sivustolla on listattu tunnetuimmat suomalaiset API-rajapinnat. Sivusto listaa suomalaisia API-rajapintoja 84 kappaletta. Palvelun tarkoituksena on helpottaa rajapintojen löytämistä aihealueittain luokiteltujen rajapintakuvausten avulla. Rajapintalistauksen lisäksi sivustolla on työkaluja ja ohjeita rajapintojen käyttöön, joiden avulla rajapintojen käyttöä pyritään helpottamaan. (APISuomi, 2015.)

Yksi tunnetuimmista suomalaisista avoimista API-rajapinnoista on Helsingin seudun liikenteen (HSL) rajapinta, jonka avulla kehittäjät pystyvät rakentamaan reaaliaikaisia joukkoliikennesovelluksia HSL:n datan päälle (Lindman et al., 2014). Toinen esimerkki yksityisen yrityksen avoimesta API-rajapinnasta on Finnkinon<sup>17</sup> XML pohjainen API-rajapinta, jonka kautta voi löytää tietoa teattereista sekä niissä pyörivistä elokuvista ja aikatauluista.

### 3.3 Avoimen datan edistäminen

Avointa dataa pyritään edistämään Suomessa monin tavoin. Suurimpien suomalaisten niin sanottujen kuutoskaupunkien Helsingin, Espoon, Vantaan, Oulun, Tampereen ja Turun yhteistyöstrategia 6Aika pyrkii edistämään avointa dataa lokaalin kaupunkitason osalta tavoitteenaan paikallisten yritysten innovaatiotoiminnan vahvistaminen. 6Aika-strategian kolme painopistettä ovat: avoimet innovaatioalustat, avoin data ja rajapinnat sekä avoin osallisuus. Strategian toimienpiteinä pyritään avaamaan tietovarantoja sekä ohjelmointirajapintoja kuutoskaupunkien järjestelmiin kenen tahansa hyödynnettäväksi. Strategian tarkoituksena on myös edistää avoimen datan hyödyntämistä toteuttamalla projekteja, joissa kehitetään avoimen datan palvelukokonaisuutta analysoimalla avattua dataa, sekä tekemällä datan hyödyntämisestä helpompaa sekä aktivoimalla yrityksiä avoimen datan hyödyntämiseen uusien sovellusten ja palvelujen aikaansaamiseksi. Strategiassa yhteistyötä yritysten kanssa pyritään kehittämään ja yritykset halutaan mukaan kertomaan omista datatarpeistaan sekä kehittämään kaupunkien kanssa yhteisiä rajapintastandardeja. Myös yrityksiä toivotaan avaavan omaa dataansa muiden käyttöön. 6Aika hanke jatkuu vuoden 2017 kesäkuun loppuun asti ja sen budjetti on 8.1 miljoonaa euroa. (6aika, 2014, Kenttä, 2015.)

Avointa dataa edistetään myös erilaisten kilpailujen avulla, joista tunnetuin Suomessa järjestettävä avoimen datan kilpailu on Apps4Finland. Ensimmäistä kertaa kilpailu järjestettiin vuonna 2009 Apps4Democracy Finland nimellä, ja se on myös Euroopan vanhin avoimen datan kilpailu. Paikallisen tason kilpailuja järjestetään myös, kuten Apps4Pirkanmaa, jota on järjestetty vuodesta 2013 lähtien. Viimeisimpinä lisäyksinä paikallisiin kilpailuihin ovat Apps4Pohjois-Pohjanmaa sekä Apps4Lounais-Suomi, jotka molemmat järjestettiin ensimmäistä kertaa vuonna 2014. Tällaisten kilpailujen konkreettiset ja toimivat sovellukset sekä hyvät datan käyttöideat kannustavat julkishallintoja avaamaan lisää tietovarantojaan. (Eteläaho, 2014.) Ensimmäisten Suomessa järjestettyjen avoimen datan kilpailujen pysyvimpänä hyötynä voidaan nähdä suomalaisen avoimen datan ekosysteemin syntyminen, jonka kehittämiseen kilpailuihin osallistuneet hallinnon organisaatiot ja kilpailijat ovat vaikuttaneet. (Poikola et al., 2010.)

---

<sup>17</sup> <http://www.finnkino.fi/xml>

Eteläaho (2014) mukaan avoimen datan kilpailutöiden perusteella voidaan olettaa, että avointen datojen käyttö kaupallisissa sovelluksissa ei ole vielä kovin yleistä. Syynä tähän Eteläaho (2014) toteaa, että dataa osattaisiin kyllä hyödyntää, mikäli niihin olisi helppo pääsy rajapintojen välityksellä. Myöskään yksityiset yritykset Suomessa eivät ole tottuneita datan avaajia, ja merkittävimmät yritysten avaamat tietoaaineistot liittyvät datajournalismiin.

Suomessa järjestetään myös hackathon-tyyppisiä avoimen datan tapahtumia, jossa joukko ihmisiä kokoontuu suunnittelemaan ja toteuttamaan yhteistyössä avoimeen dataan pohjautuvia sovelluksia. Pääasiallisena tavoitteena näillä tapahtumilla on hyödyllisten palveluiden ja sovellusten luominen. Ne mahdollistavat kuitenkin myös eri taustoista tulevien ammattilaisten kohtaamisen sekä uusien kontaktien solmimisen. Esimerkkejä tällaisista hackathon tapahtumista ovat muun muassa avoimiin kulttuuriaineistoihin keskittyvä *Hack4FI – Hack your heritage* sekä pääkaupunkiseudun avoimeen dataan keskittyvä *Forum Virium Helsinki – The Open Data Day Hackathon* tapahtuma. (Hack4Fi, 2015; forumvirium, 2015.)

## 4. Pohdinta ja johtopäätökset

Tutkimuksen tavoitteena oli selvittää edellytyksiä avoimen datan innovaatio toiminnalle sekä näiden edellytysten toteutuminen Suomessa. Avoimen datan lähteet ovat useimmissa tapauksissa julkishallintojen tuottamat tietovarannot, joten julkishallintojen avoin data on nähtävissä innovointipotentialin kannalta tärkeimpänä (Jaakkola et al., 2014a; Gurin, 2013a). Luvun 2.2 perusteella voidaan todeta, että avoimen datan potentiaali ei riipu kuitenkaan sen lähteestä, vaan myös yksityisten yritysten data-aineistot sekä API-rajapinnat ovat potentiaalisia lähteitä avoimeen dataan perustuville innovatiivisille sovelluksille.

Minimivaatimus avoimen datan hyödyntämiselle on esitysmuodon avoimuus sekä kone-luettavuus (Poikola et al., 2010). Viiden tähden luokittelun perusteella (luku 2.3) voidaan todeta, että ihannetilanteessa avoin data on linkattavissa sitä käyttävään sovellukseen sellaisessa koneluettavassa formaatissa, että se ei vaadi ylimääräistä prosessointia ennen kuin sitä voidaan hyödyntää. Muita vaatimuksia avoimelle datalle ovat dataan liittyvä lisensointi, jonka tulee sallia datan vapaa käyttö ilman rajoitteita.

Lukujen 2.3 sekä 2.4 perusteella voidaan todeta, että avoimen datan tulisi olla helposti saatavilla, teknisesti laadukasta, sekä kehittäjien kannalta mielenkiintoista. Jotta avoin data olisi kehittäjien kannalta mielenkiintoista, tulisi sen olla reaaliaikaista muuttuvaa dataa, joka on myös luotettavaa ja sen jatkuvuus on taattua. Datan tulisi olla myös helposti integroitavissa muiden data-aineistojen kanssa, jotta saavutettaisiin uusia innovatiivisia datan käyttötapoja.

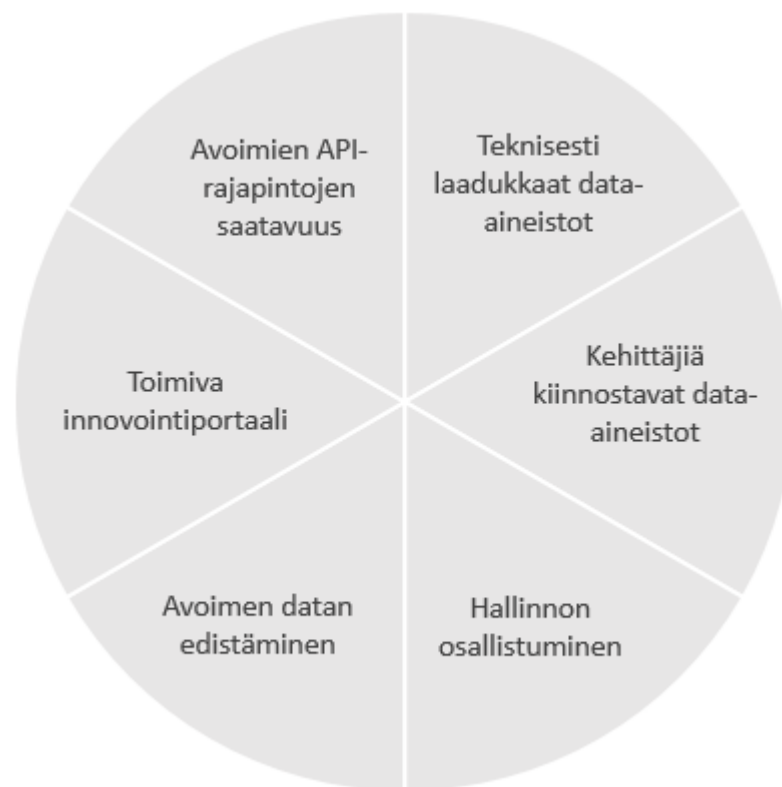
Innovatiivisten sovellusten kannalta oleellimmat kehittäjiä kiinnostavat data-aineistot ovat luvun 2.2 perusteella kaikenlainen paikkatietoon perustuva data, reaaliaikainen joukkoliikenteen data, rikostilastot sekä kansainväliset pörssidatat. Historiatietoon perustuvaa tilastodataa ei niinkään arvosteta kehittäjien keskuudessa innovatiivisten sovellusten tietoinestoina. Lokaalin avoimen datan tarjonta kiinnostaa erityisesti kehittäjiä ja kysyntää lokaaliin dataan perustuville sovelluksille on laajasti. Lokaaleihin data-aineistoihin perustuvat sovellukset myös rohkaisevat kaupunkeja avaamaan lisää tietoinestojaan. (Desouza & Bhagwatar, 2012; Jaakkola et. al., 2014b.)

Avoimien API-rajapintojen merkitys uusien innovatiivisten sovelluksien ja palvelujen mahdollistajana, voidaan pitää tärkeänä, sillä ne mahdollistavat helpon linkityksen sovelluksen ja datan välille (Dietrich et al., 2012). Avoimet API-rajapinnat ovat myös kehittäjien kannalta kiinnostavia, koska ne tarjoavat suoran pääsyn reaaliaikaisesti päivittyvään ja ajantasaiseen dataan (Peres Núñez, 2014). API-rajapinnat toimivat myös katalyyttinä innovaatioille, sillä ne voivat edesauttaa aktiivisen kehittäjäyhteisön synty-miseen niiden ympärille (Weiss & Gangadharan, 2010).

Avoimien datojen julkaisukanavat ovat näkyvimmat keinot edistää avointa dataa sekä sen saatavuutta. Julkaisukanavat eli avoimen datan portaalit voivat helpottaa avoimen datan löytämistä sekä niiden laadun arvioimista. Ihannetilanteessa tällaiset portaalit myös opastavat avoimen datan käyttöön. Portaalissa esillä olevat avoimesta datasta toteutetut sovellukset myös inspiroivat kehittäjiä luomaan uusia innovatiivisia sovelluksia avoimesta datasta. (Chan, 2013; O' Reilly, 2011; Millard, 2011.)

Avointa dataa pyritään edistämään myös erilaisten kilpailujen sekä hackathon-tapahtumien muodossa, jotka levittävät avoimen datan sanomaa, edesauttavat innovatiivisten sovellusten syntymistä, sekä luovat pohjaa avoimen datan ekosysteemin kehittymiselle (Chan, 2013; Desouza & Bhagwatar, 2012; Poikola et al., 2010).

Julkishallintojen osallistuva rooli on myös tärkeä avoimen datan edistämisessä. Dialogia datan julkaisijoiden ja käyttäjien välillä tarvitaan, jotta avatuista data-aineistoista voidaan antaa palautetta ja voidaan tunnistaa tärkeät data-aineistot, jotka tulisi seuraavaksi avata. (Poikola et al., 2010.) Julkishallinto voi toimia myös kehittäjien kanssa yhteistyössä sovelluksien rakentamisessa, jolloin julkishallinnoilla on parempi mahdollisuus vaikuttaa sovelluksien laatuun ja luonteeseen (Chan 2013). Myös lokaaleilla kaupunkitason toimilla kaupunki voi luoda aktiivisen kehittäjäyhteisön rakentamaan innovatiivisia sovelluksia kaupunkilaisten käyttöön (Walravens & Ballon, 2011; O' Reilly, 2011).



Kuva 4. Osa alueet avoimen datan innovoinnin mahdollistajina.

Tutkimuksen perusteella löydetty osa-alueet avoimen datan innovoinnin mahdollistajina ovat esitetty kuvassa 4. Perusedellytyksenä avoimen datan hyödyntämiselle on *teknisesti laadukkaat data-aineistot*, jotka ovat koneluettavissa avoimissa tiedostoformaateissa sekä julkaistu avoimella lisenssillä. Data aineistojen teknisen laadun lisäksi, tulisi data aineistojen olla *kehittäjien kannalta kiinnostavaa*, kuten paikkatietoon pohjautuvaa, reaaliaikaista tai lokaalia dataa. *Avoimet API-rajapinnat* houkuttelevat kehittäjiä hyödyntämään avointa dataa ja antavat siihen kehittäjille helpon keinon. *Avoimen datan edistäminen* kilpailujen ja tapahtumien muodossa tuo avointa dataa kehittäjien tietoisuuteen, mutta se myös auttaa kehittämään avoimen datan ekosysteemiä. *Julkishallintojen osallistuminen* datan avauksissa ja avoimen datan edistämisessä auttaa parantamaan avoimen datan laatua sekä mahdollistaa yhteistyön datan julkaisijoiden ja hyödyntäjien välillä. Hyvin *toimiva innovointiportaali* auttaa avoimen datan

hyödyntäjää löytämään mielenkiintoisia data-aineistoja ja ohjeistuksia, sekä inspiroi kehittäjiä luomaan uusia sovelluksia avoimista datoista.

Näiden osa-alueiden toteutuminen Suomessa on kohtalaisen hyvä. Avointa dataa edistetään hyvin laajasti ja hallituksen periaatepäätös avoimen datan edistämisestä sekä avoimen hallinnon toimintasuunnitelma asettavat avoimen datan tulevaisuudelle korkeat tavoitteet (Avoin hallinto, 2013; Koski et al., 2012; Kiuru et al., 2012). Tavoitteet avoimen datan suhteen ovat korkealla, ja aika näyttää miten asetetut tavoitteet toimivat käytännössä. Avoimen datan hyödyntäminen on kuitenkin vielä varsin harvinaista suomalaisissa yrityksissä, mutta sen potentiaali on laajasti tunnustettu (Kiuru et al., 2012; Immonen et al., 2014b).

Suomalaiset julkishallinnot tuottavat laadukkaita tietoaineistoja ja ovat kohtuullisen nopeita julkaisemaan niitä (Lindman et al., 2014). HRI- sekä avoindata.fi- portaalien data-aineistojen perusteella voidaan kuitenkin todeta, että koneluettavien ja avoimien tiedostomaattien mukaisia avoimia datoja ei määrällisesti ole vielä kovinkaan paljon. Voidaankin sanoa, että *teknisesti laadukkaiden data-aineistojen saatavuus* on Suomessa kohtalainen, mutta lisää laadukasta ja koneluettavaa dataa tarvitaan.

Suomesta löytyy kuitenkin kehittäjiä kiinnostavia data-aineistoja, kuten koko maan kattavat kartta-aineistot, sekä reaaliaikaista julkisen liikenteen paikkatietoa. Suurin osa saatavilla olevista avoimista datoista on kuitenkin sellaista, etteivät ne ole kehittäjien näkökulmasta mielenkiintoista, vaan pohjautuvat historialliseen tilastodataan (Peltola, 2014). Kehittäjiä kiinnostavan lokaalin datan saatavuus vaihtelee suuresti eri kaupunkeja vertaillessa. Etelä-Suomen kaupunkien osalta lokaalia dataa on hyvin saatavilla HRI-portaalin välityksellä. Myös Tampereen kaupungin avoimen datan portaali tarjoaa riittävästi avointa dataa lokaalien sovellusten tarpeisiin. Muissa suomen kaupungeissa avoimen lokaalin datan tarjonta on vielä hyvin vähäistä. Kuitenkin kuuden suurimman suomalaisen kaupungin yhteistyöstrategia 6Aika pyrkii edistämään lokaalin avoimen datan saatavuutta, joten tilanne lokaalin avoimen datan suhteen on vielä kehittymässä. *Kehittäjiä kiinnostavien data-aineistojen* saatavuus on Suomessa kohtalaista ja kehitystä sen suhteen on nähtävissä.

*Avoimia API-rajapintoja* Suomessa listaavan API Suomi sivuston, avoindata.fi-, sekä HRI- portaalien perusteella voidaan todeta, että avoimia API rajapintoja ei määrällisesti ole vielä kovinkaan paljon. Rajapintoja on kuitenkin saatavilla ja kehittäjiä kiinnostavista aihealueista kuten karttadatasta sekä joukkoliikennedatasta.

*Avointa dataa edistetään* Suomessa monin tavoin. Suurimpien suomalaisten kaupunkien yhteistyöstrategia 6Aika pyrkii edistämään avoimen datan laatua sekä parantamaan sen saatavuutta ja tietoisuutta avoimesta datasta (6Aika, 2014). Myös vuosittain järjestettävä Apps4Finland kilpailu sekä erilaiset hackathon-tapahtumat edistävät avointa dataa Suomessa. Ensimmäiset Apps4Finland kilpailut ovatkin luoneet hyvin pohjaa suomalaiselle avoimen datan ekosysteemille. (Eteläaho, 2014; Poikola et al., 2010.)

*Julkishallintojen osallistuminen* Suomessa näkyy siinä, miten avointa dataa pyritään edistämään. Avoimia tietoaineistoja pyritään tuomaan aktiivisesti saataville ja olemassa olevien aineistojen laatua pyritään parantamaan. (Avoin hallinto, 2013.) 6Aika strategia on osoitus siitä, että julkishallinnot haluavat pelkän data-avauksien lisäksi olla mukana myös rakentamassa avoimeen dataan perustuvaa tulevaisuutta suomalaiseen it-teollisuuteen (6Aika, 2014).



Suomen tunnetuimmat ja suurimmat avoimen datan julkaisukanavat avoindata.fi- sekä Helsinki Region Infoshare (HRI)- portaalit perustuvat molemmat avoimen lähdekoodin CKAN-ohjelmistoon (HRI, 2015; avoindata.fi, 2015). Portaalien kohtalaisen hyvin kattavat hakuominaisuudet ovatkin siten hyvin samankaltaisia. HRI-portaalissa on esillä avoimesta datasta tehtyjä sovelluksia, toisin kuten avoindata.fi-portaalissa, joten HRI-portaali tukee paremmin avoimen datan innovaatioita inspiroimalla kehittäjiä. Socrata-ohjelmistoon perustuva avoimen datan portaali antaisi myös mahdollisuuden julkaista avoimia datoja automaattisesti luotujen rajapintojen välityksellä. Suomessa käytössä oleva CKAN pohjaisissa portaaleissa data on saatavilla vain siinä muodossa, jossa se on sinne ladattu. Paikallisen tason avoimen datan portaalit ovat vielä vaatimattomia näihin kahteen suurimpaan portaaliin verrattuna niin data-aineistojen määrän sekä portaalien toteutuksien osalta. Näin ollen *toimivan innovointiportaalin* toteutus on Suomessa vielä alussa.

Vaikka avoin data on Suomessa vielä varsin uusi asia, mahdollisuudet avoimeen dataan perustuvaan innovaatiotoimintaan ovat kohtuullisen hyvät. Kun lisää data-aineistoja saadaan avattua sekä data-aineistojen laatua parannettua, ovat mahdollisuudet yhä parempia. Lisäksi avoimen datan hyödyt ovat Suomessa hyvin tunnustettu ja halua avoimen datan edistämiseen on olemassa. Kehittäjien kiinnostuksen herättäminen, sekä kehittäjäyhteisön tukeminen vaatii vielä työtä. Avoimia API-rajapintoja tulisi saada enemmän tarjolle, ja avoimen datan portaalien tulisi tukea kehittäjiä paremmin. Lokaaliin avoimen datan edistämiseen tarkoitettu 6Aika hanke, sekä avoimen hallinnon ohjeistukset datan laadun parantamiseksi, luovat kuitenkin uskoa avoimen datan tulevaisuudelle Suomessa.

## 5. Yhteenveto

Avoim data edistää hallinnon läpinäkyvyyttä ja demokratiaa sekä tehostaa hallinnon palveluja mutta se mahdollistaa myös uusien innovatiivisten sovellusten ja palveluiden kehittämisen (Dietrich et al., 2012). Kuitenkin suuria menestystarinoita odotetaan yhä avoimen datan hyödyntämisessä (Jaakkola et al., 2014b). Tutkimuksessa etsittiin edellytykset avoimen datan innovoinnille avoimen datan hyödyntäjän näkökulmasta.

Kriteerit avoimelle datalle ovat datan vapaa saatavuus internetin välityksellä, datan uudelleenkäytettävyys ehdoilla jotka sallivat sen vapaan käytön ja yhdistelemisen muihin datoihin. Avoimen datan tulee olla myös saatavilla ilman rajoittavia lisenssiehtoja. (Dietrich et al., 2012.) Julkishallinnot ovat datan julkaisijoista suurimpia, joten niiden data-aineistot ovat tärkeimpiä innovatiivisten sovellusten lähteitä. Lokaali kaupunkidata sekä yksityisen sektorin julkaisema avoimet datat ovat myös nähtävissä tärkeinä resursseina innovatiivisille sovelluksille. (Bonina, 2013; Herzberg, 2014; Desouza & Bhagwatar, 2012.)

Avoimen datan laatu on tärkeää, jotta avointa dataa voidaan hyödyntää tehokkaasti. Koneluettavien tiedostoformaattien lisäksi avoimen datan tulisi olla myös dynaamista, luotettavaa, reaaliaikaista, jatkuvaa sekä helposti integroitavaa, jotta se olisi kiinnostavaa kehittäjien kannalta. (Immonen et al. 2014b, Jaakkola et al., 2014b; Poikola et al., 2010.) Avoimet API-rajapinnat tarjoavat ajantasaista ja reaaliaikaisesti päivittyvää dataa, mahdollistavat helpon linkityksen sitä käyttäviin sovelluksiin sekä luovat aktiivisen kehittäjäyhteisön niiden ympärille (Dietrich et al., 2012; Weiss & Gangadharan, 2010). Avoimen datan helppo saatavuus avoimen datan portaalien välityksellä ei yksistään riitä, vaan portaalien tulisi tarjota työkaluja, opastusta sekä toimia innovointialustana inspiroiden kehittäjiä luomaan innovatiivisia sovelluksia (O’Reilly, 2011; Chan, 2013). Avointa dataa edistetään myös erilaisten kilpailujen ja hackathon-tapahtumien muodossa, jotka luovat perustaa avoimen datan ekosysteemeille (Poikola et al., 2010).

Tutkimuksen perusteella voidaan todeta, että perusedellytykset avoimen datan innovoinnin mahdollistajina ovat teknisesti laadukkaiden sekä kehittäjien kannalta kiinnostavien data-aineistojen saatavuus. Avoimien API-rajapintojen saatavuus myös toimii katalyyttinä innovatiivisille sovelluksille. Avoimen datan helpon saatavuuden edistämiseksi sekä kehittäjien aktivointiin tarvitaan toimiva innovointiportaali, joka tarjoaa helpon pääsyn avoimiin data-aineistoihin sekä työkaluja ja opastusta avoimen datan hyödyntämiseen. Lisäksi avointa dataa tulisi edistää erilaisten kilpailujen ja tapahtumien muodossa, jotta avoin data ja se mahdollisuudet tulisivat kehittäjien tietoisuuteen, ja auttaisivat kehittämään avoimen datan ekosysteemiä. Myös julkishallintojen osallistuva rooli datan avauksissa ja avoimen datan edistämässä on tärkeää, sillä se auttaa parantamaan avoimen datan laatua sekä mahdollistaa yhteistyön datan julkaisijoiden ja hyödyntäjien välillä.

Avoim data on Suomessa vielä uusi asia. Avointa dataa edistetään monella tapaa, ja laadukkaita data-aineistoja on saatavilla, mutta sen merkitys yrityksille ei ole vielä tehnyt läpimurtoa. Avoimen datan potentiaali on kuitenkin laajasti tunnustettu suomalaisissa yrityksissä. (Kiuru et al., 2012; Immonen et al., 2014b.) Suomessa tilanne avoimen

datan suhteen on kohtuullisen hyvä, mutta parannusta tarvitaan avoimen datan laadun, API-rajapintoihin määrän, sekä avoimen datan portaalien toiminnan osalta.

## Lähdeluettelo

- 5stardata, (2012). 5 ★ Open Data, Luettu: 6.12.2014, Saatavilla: <http://5stardata.info/>
- 6aika. (2014). Avoimet ja älykkäät palvelut, Luettu: 10.11.2014. Saatavilla: <http://www.6aika.fi>
- ApiSuomi. (s.d.). Api:Suomi, Suomen rajapinnat yhdestä paikasta, Luettu: 10.02.2015, Saatavilla: <http://apisuomi.fi/>
- Avoim hallinto (2013), Avoim hallinto ja osallisuuden edistäminen Suomessa. Luettu: 31.01.2015, Saatavilla: [http://www.vm.fi/vm/fi/04\\_julkaisut\\_ja\\_asiakirjat/03\\_muut\\_asiakirjat/Avoim\\_hallinto\\_toimintasuunnitelma.pdf](http://www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/03_muut_asiakirjat/Avoim_hallinto_toimintasuunnitelma.pdf)
- Aitamurto, T., & Lewis, S. C. (2011). Open APIs and news organizations: A study of open innovation in online journalism. In International Symposium on Online Journalism, Austin, TX, April (Vol. 1).
- Bonina, C. M. (2013). New business models and the value of open data: Definitions, challenges and opportunities. London School of Economics and Political Science. Department of Management
- Boswarva, O. (2014) it's time for the private sector to release some open data too. Luettu: 22.12.2014, Saatavilla: <http://www.statlife.org.uk/opinion/1778-it-s-time-for-the-private-sector-to-release-some-open-data->
- Chan, C. M. L. (2013). From open data to open innovation strategies: Creating E-services using open government data. Proceedings of the 46th Hawaii International Conference on System Sciences (HICSS), 1890-1899. doi:10.1109/HICSS.2013.236
- Chan, C. M., Lau, Y., & Pan, S. L. (2008). E-government implementation: A macro analysis of Singapore's e-government initiatives. *Government Information Quarterly*, 25(2), 239-255.
- CKAN, (s.d.). CKAN The open source data portal software, Luettu: 11.02.2015, Saatavilla: <http://ckan.org/>
- Davies, T. (2013). Open Data Barometer: 2013 Global Report. World Wide Web Foundation and Open Data Institute. Luettu: 12.12.2014. Saatavilla: <http://www.opendataresearch.org/dl/odb2013/Open-Data-Barometer-2013-Global-Report.pdf>.
- Desouza, K. C. & Bhagwatwar, A. (2012). Citizen apps to solve complex urban problems. *Journal of Urban Technology*, 19(3), 107-136.
- Dietrich, D., Gray, J., McNamara, T., Poikola, A., Pollock, P., Tait, J., & Zijlstra, T. (2012). Open data handbook. Luettu: 10.11.2014. Saatavilla: <http://opendatahandbook.org>.

- EC (2003), Directive 2003/98/EC (Public sector information (PSI) Directive), Directive on the re-use of public sector information. EUR-Lex Europa. Luettu: 20.12.2014, Saatavilla: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:345:0090:0096:EN:PDF>
- EC (2007), Directive 2007/2/EC (INSPIRE Directive). Directive establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Luettu: 20.12.2014, Saatavilla: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF>
- Eteläaho, A. (2014) Analyysi avoimen datan innovaatiokilpailun tunnustusta saaneista kilpailutöistä 2010-2013. Luettu: 31.01.2015, Saatavilla: [https://www.tut.fi/avaras/?page\\_id=991](https://www.tut.fi/avaras/?page_id=991)
- Ferro, E. & Osella, M. (2013). Eight business model archetypes for PSI re-use. “Open Data on the Web” Workshop, Google Campus, London
- forumvirium, (2015). Open Data Day Helsinki Hackathon, Luettu: 12.02.2015, Saatavilla: <http://www.forumvirium.fi/en/event/open-data-day-helsinki-hackathon>
- Gurin, J. (2013a). Big Data vs Open Data – Mapping it out, Luettu: 15.12.2014, Saatavilla: <http://www.opendatanow.com/2013/11/new-big-data-vs-open-data-mapping-it-out/>
- Gurin, J. (2013b). Data and Science: Masters of the Zooniverse, Luettu: 15.12.2014, Saatavilla: <http://www.opendatanow.com/2013/10/new-masters-of-the-zooniverse/>
- Gurin, J. (2014). Driving Innovation with Open Data. The Future of Data Driven Innovation, p. 55 – 65. The U.S. Chamber of Commerce Foundation.
- Gurstein, M. B. (2011). Open data: Empowering the empowered or effective data use for everyone? First Monday, Volume 16, Number 2 - 7 February 2011, Luettu: 10.11.2014. Saatavilla: <http://firstmonday.org/ojs/index.php/fm/article/view/3316/2764#16a>
- Hack4Fi, (2015). Hack4FI – Hack your heritage, Luettu: 12.02.2015, Saatavilla: <http://hack4.fi/>
- Herzberg, B. (2014). The Next Frontier for Open Data: An Open Private Sector, Luettu: 22.12.2014, Saatavilla: <http://blogs.worldbank.org/voices/next-frontier-open-data-open-private-sector>
- HRI, (2015). Helsinki Region Infoshare – avointa dataa pääkaupunkiseudulta vapaasti hyödynnettäväksi, Luettu: 4.2.2015, Saatavilla: <http://www.hri.fi/fi/>
- Immonen, A., Palviainen, M. & Ovaska, E. (2014a). Requirements of an open data based business ecosystem. Access, IEEE, 2, 88-103. doi:10.1109/ACCESS.2014.2302872
- Immonen, A., Palviainen, M., & Ovaska, E. (2014b). Towards open data based business: Survey on usage of open data in digital services. International Journal of Research in Business and Technology, 4(1), 286-295.

- Jaakkola, H., Mäkinen, T., Henno, J. & Mäkelä, J. (2014a). Open. 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 608-615. doi:10.1109/MIPRO.2014.6859640
- Jaakkola, H., Mäkinen, T. & Eteläaho, A. (2014b). Open data: Opportunities and challenges. Proceedings of the 15th International Conference on Computer Systems and Technologies, Ruse, Bulgaria. 25-39. doi:10.1145/2659532.2659594
- Kenttä., R. (2015). Oulu ja kumppanit avaavat dataansa. Forum24 15.01.2015
- Kiuru P., Mäkelä J., & Huvio P. (2012). Avoimen julkisen tiedon hyödyntämisen potentiaalista suomalaisissa yrityksissä. Aalto-yliopiston kauppakorkeakoulun pienyrityskeskus. Aalto-yliopiston insinööritieteiden korkeakoulun maankäyttötieteiden laitos. Digires RY.
- Korn, N., & Oppenheim, C. (2011). Licensing Open Data: A Practical Guide. Luettu: 04.01.2015, Saatavilla: [http://discovery.ac.uk/files/pdf/Licensing\\_Open\\_Data\\_A\\_Practical\\_Guide.pdf](http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide.pdf)
- Koski, H., Kiuru, P., Mäkelä, J., & Salokannel, M. Julkinen tieto käyttöön. Luettu: 31.01.2015, Saatavilla: <http://www.etla.fi/wp-content/uploads/2012/09/dp1276.pdf>
- Lakomaa, E. & Kallberg, J. (2013). Open data as a foundation for innovation: The enabling effect of free public sector information for entrepreneurs. Access, IEEE, 1, 558-563. doi:10.1109/ACCESS.2013.2279164
- Lindman, J., Kinnari, T. & Rossi, M. (2014). Industrial open data: Case studies of early open data entrepreneurs. Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS), 739-748. doi:10.1109/HICSS.2014.99
- Manyika, J., Chui, M., Groves, P., Farrel, D., Van Kuiken, S. & Doshi, E. A. (2013). Open data: Unlocking innovation and performance with liquid information. McKinsey Global Institute.
- Millard, J. (2013). ICT-enabled public sector innovation: trends and prospects. In Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance (pp. 77-86). ACM.
- O'Reilly, T. (2011). Government as a platform. Innovations, 6(1), 13-40.
- Peltola. V. (2014). Tämän takia ei ole avoimen datan startuppeja. Luettu: 7.2.2015, Saatavilla: <http://www.taloussanommat.fi/kumppaniblogit/2014/11/17/taman-takia-ei-ole-avoimen-datan-startuppeja/201415884/322>
- Peres Núñez, W. (2014). Big data and open data as sustainability tools: A working paper prepared by the Economic Commission for Latin America and the Caribbean.
- Poikola, A., Kola, P. & Hintikka, K. A. (2010). Julkinen data–johdatus tietovarantojen avaamiseen. Helsinki: Edita. Luettu: 10.02.2015, Saatavilla: [https://www.lvm.fi/docs/fi/964902\\_DLFE-10617.pdf](https://www.lvm.fi/docs/fi/964902_DLFE-10617.pdf)
- Rohunen, A., Markkula, J., Heikkilä, M. & Heikkiä, J. (2014). Open traffic data for future service innovation: addressing the privacy challenges of driving data. Journal of Theoretical and Applied Electronic Commerce Research, 9(3), 71-89.

- Salminen, J. (2013). Avoin tieto tuo miljardeja ja uudistaa demokratiaa – näin se tehdään. Suomen kuvalehti. Luettu: 7.2.2015, Saatavilla: <http://suomenkuvalehti.fi/jutut/kotimaa/avoin-tieto-tuo-miljardeja-ja-uudistaa-demokratiaa-nain-se-tehdaan/>
- SOCRATA, (2015). Socrata Open Data Portal, Luettu: 11.02.2015, Saatavilla: <http://www.socrata.com/products/open-data-portal/>
- The Five Stars of Open Data, (2010), Luettu: 6.12.2014, Saatavilla: [http://www.epsiplatform.eu/sites/default/files/The%205%20stars%20of%20Open%20Data\\_MdV\\_PR2.pdf](http://www.epsiplatform.eu/sites/default/files/The%205%20stars%20of%20Open%20Data_MdV_PR2.pdf)
- Ubaldi, B. (2013), Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives, OECD Working Papers on Public Governance, No. 22, OECD, Saatavilla: <http://dx.doi.org/10.1787/5k46bj4f03s7-en>
- Veijalainen, A., (2012), Data avautuu – Maanmittauslaitos näyttää tietä. Esri Finland Avoin data extra s.4-5. Esri Finland Oy. Luettu: 7.2.2015, Saatavilla: [http://www.esri.fi/midcom-serveattachmentguid-1e1b6caff53a732b6ca11e1bd52850b41011c621c62/avoindataextra\\_lowres.pdf](http://www.esri.fi/midcom-serveattachmentguid-1e1b6caff53a732b6ca11e1bd52850b41011c621c62/avoindataextra_lowres.pdf)
- Walravens, N., & Ballon, P. (2011). The City as a Platform: Exploring the Potential Role (s) of the City in Mobile Service Provision through a Mobile Service Platform Typology. In Tenth International Conference on Mobile Business (ICMB), (pp. 60-67). IEEE.
- Weiss, M., & Gangadharan, G. R. (2010). Modeling the mashup ecosystem: structure and growth. *R&d Management*, 40(1), 40-49.
- Zeleti, F. A., Ojo, A., & Curry, E. (2014). Emerging business models for the open data industry: characterization and analysis. In Proceedings of the 15th Annual International Conference on Digital Government Research (pp. 215-226). ACM.
- Zuiderwijk, A., Janssen, M., & Davis, C. (2014). Innovation with open data: Essential elements of open data ecosystems. *Information Polity*, 19(1), 17-33.