

Sirpa Peltari

AVOIN DATA

AVOIN DATA

Sirpa Peltari
Opinnäytetyö
Syksy 2014
Tietojenkäsittely
Oulun ammattikorkeakoulu

TIIVISTELMÄ

Oulun ammattikorkeakoulu
Tietojenkäsittely

Tekijä(t): Sirpa Peltari

Opinnäytetyön nimi: Avoin data

Työn ohjaaja: Ritva Virkkala

Työn valmistumislukukausi- ja vuosi: Syksy 2014

Sivumäärä: 33

Opinnäytetyön aihe syntyi toimeksiantajan esittämästä kysymyksestä, mitä tietojenkäsittelyn ammattilaisten tulisi tietää avoimesta datasta. Tietojenkäsittelyn ja tietotekniikan osaajat ovat keskeisessä roolissa avoimen datan jalostamisessa sovelluksiksi, analyyseiksi ja visuaaliseksi informaatioksi, joten aihe on ajankohtainen ja otollinen tietojenkäsittelyn ammattialalle. Tuoreudestaan johtuen aiheeseen liittyvää opetus- ja tutkimusmateriaalia ei vielä alkuvuodesta 2014 ollut löydettävissä.

Tutkimustehtävän tarkoituksiksi muotoutui selvittää, mitä tarkoitetaan avoimella datalla ja tietovarantojen avaamisella, sekä mitä vaatimuksia tietovarantojen avaamisessa on otettava huomioon. Lisäksi haluttiin kartoittaa, millaisia tietovarantoja Suomessa on avattu. Näkökulma tutkimukseen on empiirinen uuden ilmiön yleisselvitys. Tutkimuksen tietoperusta kerättiin aiheeseen liittyvistä julkaisuista, artikkeleista, muistioista ja raporteista.

Avoimen datan määritelmä pitää sisällään useita käsitteitä, kuten laillinen ja vapaa uudelleenkäytettävyys, maksuttomuus, koneluettavuus ja löydettävyys. Data-aineisto on avointa, kun kuka tahansa voi sitä vapaasti käyttää, muokata ja uudelleen jakaa. Julkisten tietovarantojen vapaan uudelleenkäytön katsotaan edistävän muun muassa demokratiaa ja hallinnon läpinäkyvyyttä, hallinnon palveluiden tehokkuuden ja tuottavuuden paranemista, kansalaisten osallistumismahdollisuutta sekä uusien tuotteiden ja palveluiden syntymistä.

Tietovarantojen avaamisessa on huomioitava, että julkaistavat data-aineistot ovat lainsäädännön puitteissa julkisia, ja että aineiston hyödyntäjillä on laillinen oikeus aineiston uudelleenkäyttöön. Tietovarantojen julkisuutta määrittävät muun muassa julkisen hallinnon asiakirjojen ja muiden tietoaineistojen julkisuudesta säättävä julkisuuslaki, henkilötietojen suojasta ja henkilötietojen käsittelystä säättävä henkilötietolaki, sekä henkilötietojen käsittelyä koskevat erityislait ja -säännökset.

Tietovarantojen avaamisella on edistetty arkea helpottavien ja hallinnon läpinäkyvyyttä lisäävien palveluiden syntymistä. Avointa dataa on hyödynnetty erilaisissa verkko- ja mobiilisovelluksissa sekä visualisoinneissa ja analyyseissä. Kuinka paljon kokonaan uutta liiketoimintaa ja innovatiivista toimintaa avoimen datan ympärille kehittyy, nähdään vasta, kun tietovarantojen avaamisesta saadaan enemmän kokemusta ja tutkimustietoa.

Asiasanat: avoin data, julkinen data, maksuttomuus, koneluettavuus, löydettävyys, uudelleenkäytettävyys

ABSTRACT

Oulu University of Applied Sciences
Degree programme in Business Information Systems

Author(s): Sirpa Pelttari

Title of thesis: Open data

Supervisor(s): Ritva Virkkala

Term and year when the thesis was submitted: Autumn 2014 Number of pages: 33

The research task of this thesis was to study new phenomena of open data. The thesis discusses the why, what and how of open data – why to go open, what open is, and how to open data. In addition, it examines the Finnish data resources published according to as open data principles. The theoretical framework of a research project relates to the related publications, articles, memos, and reports.

Open data definition includes a number of notions, such as legal and free reusability, non-chargeability, machine-readability and accessibility. Open data is data that can be freely used, reused and redistributed by anyone. Open data needs to be technically and legally open, and specifically, the data needs to be available in a machine-readable format. When planning to open data, one should find out which of the data, held by an organization are public. This should be done in accordance with legislation and policies. These include, among others, the Freedom of Information Act, data protection legislation and the Copyright Act.

Opening public data resources contributes to the goals like the transparency of democracy and administration, creation of new innovative products and services, and the increase of efficiency within the government. Open data is a relevant and topical issue in computing science, as the programmers have an important role in transforming open data into visualization and other illustration, and in creating new internet services and applications.

Keywords: open data, public data, non-chargeability, machine-readability, accessibility, reusability

SISÄLLYS

SISÄLLYS	5
1 JOHDANTO	6
2 AVOIMEN DATAN AJATUSMALLI	7
2.1 Data käsitteenä	7
2.2 Avoin julkinen tieto eli avoin data	9
2.3 Avoin data ilmiönä	10
3 DATAN AVOIMUUDEN KRITTEERIT	13
3.1 Julkisuus	13
3.2 Maksuttomuus	15
3.3 Koneluettavuus	16
3.4 Löydettävyys	17
3.5 Uudelleenkäytettävyys	18
4 DATAN JULKAISUPROSESSI	20
5 DATAN AVAUKSIA SUOMESSA	23
5.1 Avoindata.fi-portaali	23
5.2 Muita avoimen datan lähteitä	24
5.3 Esimerkkejä avoimen datan hyödyntämisestä	26
6 JOHTOPÄÄTÖKSET	28
7 POHDINTA	30
LÄHTEET	31

1 JOHDANTO

Jo muutaman vuoden ajan avoin data on ollut nouseva trendi, ja avoimen datan ajatusmalli eli julkisten tietovarantojen avaaminen kansalaisten ja yritysten vapaaseen ja maksuttomaan käyttöön on saanut nopeasti jalansijaa Suomessa. Julkisen hallinnon tietovarantojen avaamista on edistetty erilaisilla valtionhallinnon periaatepäätöksillä ja ohjelmilla, ja sekä julkisia että yksityisiä toimijoita kannustetaan datan julkaisemiseen ja hyödyntämiseen järjestämällä erilaisia tapahtumia ja innovaatiokilpailuja. Avaamalla julkisia tietovarantoja kansalaisten ja yritysten käyttöön uudelleen hyödynnettäväksi tavoitellaan muun muassa uusien tuotteiden ja palveluiden ja sitä myöten uuden liiketoiminnan ja uusien työpaikkojen syntymistä. Lisäksi tietovarantojen avaamisen katsotaan edistävän kansalaisten osallistumismahdollisuutta, demokratiaa ja hallinnon läpinäkyvyyttä sekä hallinnon palveluiden tehokkuuden ja tuottavuuden paranemista.

Tämän opinnäytetyön aihe syntyi toimeksiantajan, lehtori Sinikka Viinikan, esittämästä kysymyksestä, mitä tietojenkäsittelyn ammattilaisten tulisi tietää avoimesta datasta. Tietojenkäsittelyn ja tietotekniikan osaajat ovat keskeisessä roolissa avoimen datan jalostamisessa sovelluksiksi, analyyseiksi ja visuaaliseksi informaatioksi, joten aihe on ajankohtainen ja otollinen tietojenkäsittelyn ammattialalle. Tuoreudestaan johtuen aiheeseen liittyvää opetus- ja tutkimusmateriaalia ei vielä alkuvuodesta 2014 ollut löydettävissä. Tutkimustehtävän tarkoituksiksi muotoutui selvittää, mitä tarkoitetaan avoimella datalla ja tietovarantojen avaamisella, sekä mitä vaatimuksia tietovarantojen avaamisessa on otettava huomioon. Lisäksi haluttiin kartoittaa, millaisia tietovarantoja Suomessa on avattu.

Näkökulma tutkimukseen on empiirinen uuden ilmiön yleisselvitys. Tutkimuksen tietoperusta kerättiin aiheeseen liittyvistä julkaisuista, artikkeleista, muistioista ja raporteista. Opinnäytetyössä kuvataan avoimen datan ajatusmalli ja siihen liittyvät käsitteet ja määritelmät. Aiheeseen johdatellaan selittämällä käsite data, joka korostaa tiedon tarjoamista digitaalisessa ohjelmallisesti käsiteltävissä muodossa. Tietovarantojen avoimuuden arvioimiseksi on luotu avoimen datan kriteeristö, joka pitää sisällään useita käsitteitä, kuten julkisuus, maksuttomuus, koneluetavuus, löydettävyyss ja uudelleenkäytettävyyss. Näitä käsitteitä tarkastellaan yksityiskohtaisemmin datan avoimuuden näkökulmasta. Tietovarantojen avaamisessa huomioitavia vaatimuksia käydään läpi datan julkaisuprosessia mukaillen.

2 AVOIMEN DATAN AJATUSMALLI

Julkisella hallinnolla on hallussaan laajoja tietovarantoja, joita tuotetaan ensisijaisesti hallinnollisen päätöksenteon sekä lakien toimeenpanon ja tulkinnan tueksi (Valtiotason arkkitehtuurit -hanke 2010, 8). Julkisen hallinnon tuottama tieto on julkista, jollei julkisuutta ole laissa erikseen rajoitettu. Julkiseen tietoon on kaikilla tiedonsaantioikeus, mutta kansalaisia, yksityistä sektoria, järjestöjä ja tutkimuslaitoksia on perinteisesti pidetty ainoastaan tiedon loppukäyttäjinä. Avoimen datan ajatusmallissa tietovarannoissa oleva arvokas raakadata halutaan hyödynnettäväksi mahdollisimman laajasti niin, että dataa voidaan vapaasti ja maksuttomasti käyttää ja jalostaa omiin tarkoituksiin tasavertaisesti julkisen hallinnon kanssa. (Poikola, Kola & Hintikka 2010, 11–12, 15, 47.)

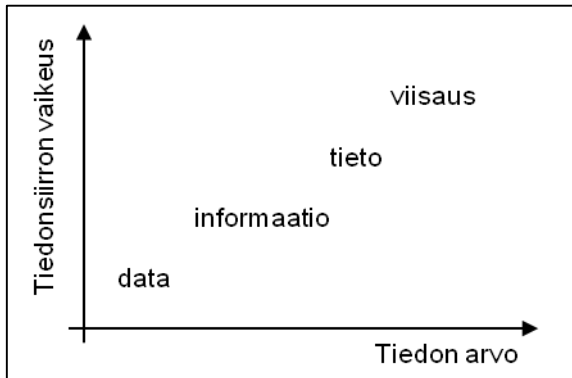
Digitaalinen tiedonvälitys on arkipäiväistynyt nopeasti, ja internet mahdollistaa tehokkaan ja taloudellisen tavan jakaa dataa julkishyödykkeenä. Avointa dataa voidaan hyödyntää esimerkiksi internet- ja mobiilisovelluksissa, jotka visualisoivat ja koostavat tietoa eri lähteistä ja tuovat arkea helpottavia tai päätöksenteon läpinäkyvyyttä lisääviä palveluja loppukäyttäjälle. Koulutuskäytössä dataa voidaan hyödyntää asioiden havainnollistamiseen, ja tutkimus- ja tuotekehitysorganisaatioissa datan helppo saatavuus luo edellytyksiä laadukkaalle tutkimustyölle ja uusille innovaatioille. (Poikola ym. 2010, 22, 55.)

2.1 Data käsitteenä

Suomen arkikielessä sanat tieto, informaatio ja data mielletään usein samaa tarkoittavaksi. Myös tiedon avoimuuden ja tietovarantojen uudelleenkäytön yhteydessä puhutaan usein sekä avoimesta datasta että avoimesta tiedosta. Termi data kuitenkin korostaa tiedon tarjoamista raaka-aineena: digitaalisessa, koneellisesti käsiteltävässä muodossa. (Poikola ym. 2010, 13–14.)

Tietotekniikan näkökulmasta eri tiedon muodot – data, informaatio, tieto, viisaus – voidaan ajatella tietosisällön käsittelyn automatisoinnin tasoina. Data (engl. data) on tiedon raaka-ainetta, merkkejä joilla ei ole merkitystä ilman tulkintaa. Raakadataa on helppo hallita ja siirtää paikasta toiseen merkitsemättä datan sisältöä. Data muuttuu informaatioksi (engl. information), kun se tulkitaan eli vastaanottaja on antanut sille merkityksen. Väline informaation välittämiseksi voi olla esimerkiksi raakadatan sisällön selittävä teksti tai raportti. Tietoteknisin keinoin informaation sisältävien dokument-

tien siirto on periaatteessa helppoa, mutta tietosisältöjen hallinta vaikeutuu. Tulkitusta ja sisäistetyistä informaatiosta syntyy tietoa (engl. knowledge). Mitä laajempia tulkittuja kokonaisuuksia käsitellään, sitä vaikeampaa on tiedon siirto automatisoidusti erilaisiin käyttötarpeisiin (kuvio 1). (Kaario & Peltola 2008, 6–7.)



KUVIO 1. Tiedon merkityksen kasvaessa tiedon siirron automatisointi vaikeutuu (Kaario ym. 2008, 8)

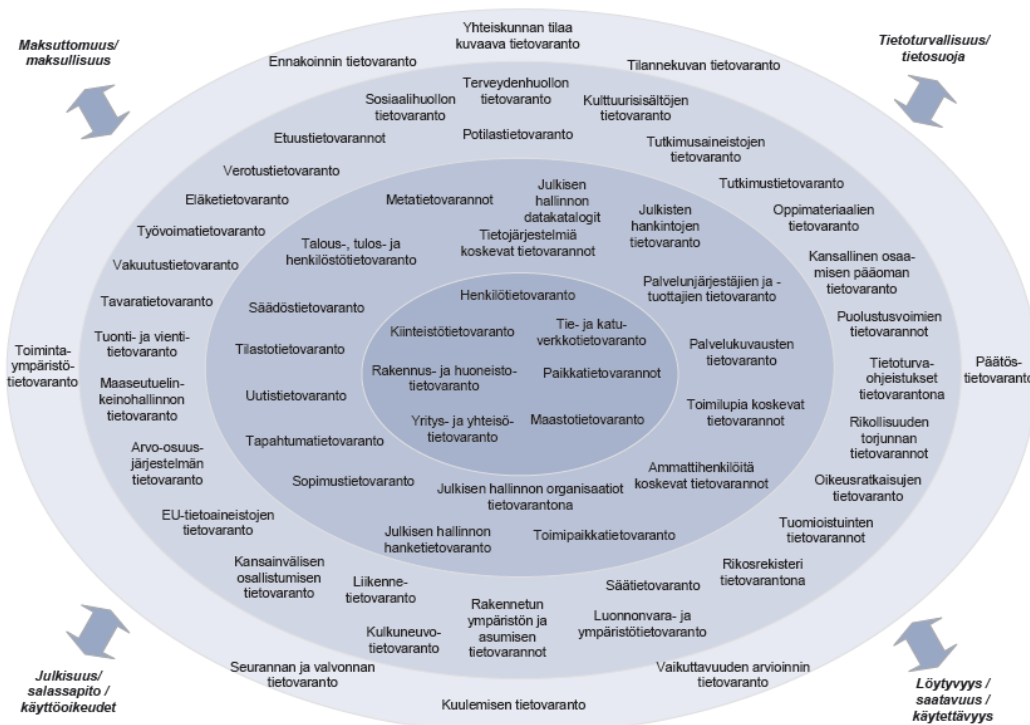
Sanastokeskus TSK:n Tietotekniikan termitalkoot (2013, hakupäivä 28.2.2014) määrittelee sanan data "tieto koneellisesti luettavassa, viestittävässä tai käsiteltävässä muodossa". Data voidaan jaotella alakäsitteisiin datan käsittelyn, avoimuuden, linkityksen, tuottajan ja määrän mukaan.

- *Raakadata* (engl. raw data) on käsittelemätöntä tai muokkaamatonta dataa. Tietyn tietojenkäsittelyvaiheen käsitelty data voi myös olla seuraavan vaiheen raakadataa, koska dataa monesti käsitellään useammassa vaiheessa.
- *Avoim data* (engl. open data) on tietyn tahon keräämää tai tuottamaa dataa, jota muut voivat maksutta ja luvallisesti hyödyntää eri tavoin uudelleen.
- *Linkitetty data* (engl. linked data) on "eri lähteissä sijaitseva rakenteinen data, jossa on lähteen sisäisiä ja lähteiden välisiä linkkejä".
- *Julkinen data* (engl. public data) on julkisen hallinnon tai muun tahon tuottamaa dataa, jonka julkaiseminen on sallittua, ja jota voidaan julkaista vapaasti käytettäväksi.
- *Massadata* (engl. big data) on dataa, jota on paljon, jota tulee nopeasti lisää, joka on muodoltaan vaihtelevaa ja jonka käsittely on vaikeaa perinteisillä tietojenkäsittelysovelluksilla.

2.2 Avoin julkinen tieto eli avoin data

Avoimen datan periaatteiden mukaisesti kaikki julkishallinnon keräämä ja tuottama data, joka on lainsäädännössä määritelty julkiseksi, voidaan avata kansalaisten ja yritysten käyttöön *uudelleen hyödynnettäväksi*. Julkinen data on valtaisa resurssi, jonka vapaan uudelleenkäytön on todettu edistävän muun muassa demokratiaa ja hallinnon läpinäkyvyyttä, hallinnon palveluiden tehokkuuden ja tuottavuuden paranemista, kansalaisten osallistumismahdollisuutta sekä uusien tuotteiden ja palveluiden syntymistä. (Dietrich, Gray, McNamara, Poikola, Pollock, Tait & Zijlstra 2012, 4–5.)

Julkisen hallinnon tietovarantoja ovat muun muassa henkilötietovaranto (väestötiedot), kiinteistö-tietovaranto, maasto- ja paikkatietovaranto, rakennus- ja huoneistotietovaranto, tie- ja katuverkko-tietovaranto sekä yritys- ja yhteisötietovaranto. Nämä perusdatavarannot ovat pääsääntöisesti yleisiä, hallinnon kohdealueesta riippumattomia tietovarantoja, ja ovat raaka-ainetta lukuisille muille tietovarannoille. Oheinen kuvio 2 kuvaa hyvin julkisen hallinnon tietovarantojen laajuutta. Tietova-rannot on jäsennelty kuvioon ympyrän keskiöstä lähtien data-informaatio-tieto/tietämys tasojen mukaan. (Valtiotason arkkitehtuurit -hanke 2010, 44–45, 48.)



KUVIO 2. Julkisen hallinnon tietovarantoja (Valtiotason arkkitehtuurit -hanke 2010, 47)

Suomessa julkisen hallinnon tietovarantojen julkisuuteen liittyy useita lakeja ja säädöksiä. Tietojen julkisuusperiaatteesta määrätään julkisuuslaissa ja tietosuojalainsäädäntö, kuten henkilötietolaki, turvaa yksityisyyden suojaa. Tekijänoikeuslaki, kilpailulainsäädäntö, kansainväliset suositukset sekä EU-direktiivit asettavat rajoituksia aineistojen edelleen hyödyntämiselle. (Poikola ym. 2010, 7, 47.)

Yksittäisen tietovarannon avoimuuden ja uudelleenkäytettävyyden arvioimiseksi on luotu avoimen datan kriteeristö (Open Definition), joka sekä auttaa tunnistamaan avoimen datan piiriin kuuluvan aineiston että varmistaa eri datalähteiden yhteentoimivuuden. *Aineisto on avointa, kun kuka tahansa voi sitä vapaasti käyttää, muokata ja uudelleen jakaa – edellyttäen korkeintaan, että datan lähde mainitaan käyttöluvasta ilmenevällä tavalla.* Tällä tarkoitetaan, että:

- Aineisto on kokonaisuudessaan saatavilla enintään kohtuullisella tuottamiskustannuksella, mieluiten maksutta. Aineisto on ladattavissa internetin kautta käyttökelpoisessa ja muokattavassa muodossa.
- Aineiston uudelleenkäyttöön, jatkojalostamiseen ja jakeluun annetaan laillinen oikeus käyttöluvalla (lisenssi).
- Aineiston käyttöä ei saa rajoittaa esimerkiksi asettamalla henkilöitä tai ryhmiä eriarvoiseen asemaan tai kieltämällä aineiston kaupallinen käyttö. (Poikola ym. 2010, 34; Dietrich ym. 2012, 6.)

2.3 Avoin data ilmiönä

Tämän raportin kirjoittamisessa hyödynnetyn lähdeaineiston perusteella vuotta 2009 voidaan pitää merkittävänä avoimen datan ekosysteemin kehittymisen kannalta Suomessa. Tuolloin järjestettiin ensimmäinen *avoimen datan innovaatiokilpailu*, jossa osallistujat saivat hyödyntää kilpailuun tarjottuja datavarantoja ja muita avoimia rajapintoja julkiseen tietoon. Linkit kilpailun tietolähteisiin olivat keskitetysti koottuna Suomi.fi -portaalissa, joka oli esiaste nykyiselle julkisen hallinnon *avoimen datan katalogille*. Alkunsa sai myös *Helsinki Region Infoshare* –hanke, joka on pääkaupunkiseudun kuntien yhteinen avoimen datan palvelu ja datan avaamisen edelläkävijä Suomessa. Elinkeinoelämän valtuuskunta EVA julkaisi raportin "Nykyäikää etsimässä - Suomen digitaalinen tulevaisuus", jossa ehdotetaan julkisen sektorin tietovarantojen avaamista maksuttomasti kaikkien käyttöön. (Poikola ym. 2010; Partanen & Upola 2013.)

Avoimen datan innovaatiokilpailusta (2009 Kansalaisosallistujan työkalut, sittemmin Apps4Finland) on kehittynyt vuosittain järjestettävä tapahtuma, jonka tarkoituksena on kannustaa julkisia toimijoita avaamaan tietojaan kansalaisten ja uusien palvelujen kehittäjien hyödynnettäviksi, ja innostaa kansalaisia ja yrityksiä kehittämään uusia tapoja hyödyntää avoimia tietovarantoja. Kilpailulla on myös merkittävä rooli avoimen datan tarjoamien etujen tietoisuuden lisäämisessä. (Forum Virium Helsinki 2014, hakupäivä 13.3.2014.) Kilpailussa palkittuja menestyneitä sovelluksia ovat muun muassa Veropuu-visualisointisovellus, Parkkinappi-pysäköintimaksusovellus ja BlindSquare-sovellus näkövammaisille.

- Veropuu-sovellus visualisoi mistä valtion, kunnan tai yksittäisen julkishallinnon organisaation tulot muodostuvat ja mihin rahat käytetään. Veropuuta hyödyntävät muun muassa yksittäiset kunnat, valtionvarainministeriö ja Suomen evankelis-luterilainen kirkko.
- Parkkinappi (nykyisin ParkMan) -pysäköintimaksusovellus hyödyntää GPS-paikkatietoja ja yhteistyökaupunkien pysäköintikarttojen tietoja maksuhyödykkeistä. Tällä hetkellä ParkMan-älypuhelinsovellus toimii Suomessa kuudessa, Ruotsissa kolmessa ja Tanskassa kahdessa kaupungissa.
- BlindSquare-sovellus antaa puhelimen ja kuulokkeiden avulla näkövammaiselle tarvittavat tiedot ympäristöstä. Sovellus käyttää hyväkseen sijaintiohjelma FourSquarea ja avointa karttapalvelua Open Street Mapia, sekä paikallisesti Helsingin kaupungin palvelukarttaa ja Helsingin seudun liikenteen avointa dataa. BlindSquare on saanut maailmanlaajuisista huomiota ja käyttäjiä yli 50 maasta. (Partanen & Upola 2013, 13–15.)

Helsinki Region Infoshare –hanke (HRI) perustettiin kokoamaan pääkaupunkiseudun kaupunkeja ja seutua koskevat tiedot yhteen palveluun, avoimesti ja maksuttomasti kenen tahansa tarvitsijan saataville. Alkuvuodesta 2011 lähtien HRI:n verkkopalveluun on kerätty muun muassa Helsingin seudun taloutta, väestöä, hyvinvointia, liikkumista ja työllisyyttä koskevia tietoaineistoja, joita voi vapaasti ladata ja hyödyntää esimerkiksi päätöksenteossa, sovellusten kehittämisessä ja tutkimus- ja kehittämistoiminnassa. (Helsinki Region Infoshare 2014, hakupäivä 14.3.2014.) Tänä päivänä HRI:n ylläpitämässä datakatalogissa on yli 1000 data-aineistoa, joista suosituimpia on Helsingin seudun liikenteen (HSL) jo vuonna 2009 avaama rajapinta joukkoliikenne- ja aikataulutietoihinsa. HSL:n jakaman avoimen tiedon pohjalta on syntynyt lukuisia matkustajainformaation saatavuutta edistäviä palveluita. Muita huomiota saaneita aineistoja ovat esimerkiksi pääkaupunkiseudun toimipisterekisteri, pääkaupunkiseudun paikkatietoaineisto asukastiedoista, rakennuskannasta ja rakennusmaavarannoista, sekä pääkaupunkiseudun kirjastojen kokoelmatiedot. (Partanen ym. 2013, 6, 16, 22–28.)

Valtiovallinnon tasolla kirjaus julkisten tietoaaineistojen avaamisen mahdollistavien periaateratkaisujen tekemisestä tehtiin hallitusohjelmaan kesäkuussa 2010 (Partanen ym. 2013, 10). Alkuvuodesta 2011 valtioneuvosto antoi periaatepäätöksen ”julkisen sektorin digitaalisten tietoaaineistojen saatavuuden parantamisesta ja uudelleenkäytön edistämisestä”, jossa esitettiin linjauksia ja toimenpiteitä ”julkisen sektorin hallinnoimien digitaalisten tietoaaineistojen saattamiseksi helposti löydettävissä ja hyödynnettävässä muodossa tietoverkkojen kautta kansalaisten, yritysten ja yhteisöjen, viranomaisten, tutkimuksen ja koulutuksen käyttöön” (Valtioneuvosto 2011, 1; Partanen ym. 2013, 39). Keväällä 2013 valtiovarainministeriö asetti avoimen tiedon ohjelman toimikaudelle 17.5.2013 - 30.6.2015. Valtiovarainministeriön johdolla käynnistettävien hankkeiden ja toimenpiteiden avulla tähdätään julkisten tietovarantojen mahdollisimman laajaan hyödyntämiseen ja tietovarantojen systemaattiseen ja mahdollisimman nopeaan avaamiseen. (Valtiovarainministeriö 2013, 2; Partanen ym. 2013, 38–39.)

Valtion laitoksista ensimmäisenä tietoaaineistojaan avasi Maanmittauslaitos. Digitaalisia maksuttomia maastotietoaaineistoja sisältävä tiedostopalvelu avattiin keväällä 2012. Ilmatieteen laitos avasi kesällä 2013 rajapinnan säähavainto- ja ennusteaineistoihinsa. Avoimen tiedon ohjelman avulla tietovarantojen avaamista jatketaan tilastoaineistoilla, yritystiedoilla, liikenne- ja viestintädatalla sekä lentokenttähavainnoilla. Osana avoimen tiedon ohjelmaa toteutetaan myös tietovarantojen löytämistä ja hyödyntämistä helpottava uusi datakatalogi sekä avoimeen dataan perustuvien palvelujen kehittämisympäristö. (Valtiovarainministeriö 2013, 2; Partanen ym. 2013, 30, 38.)

3 DATAN AVOIMUUDEN KRITTEERIT

Avoimen datan ajatusmallin keskeinen periaate on julkisin varoin tuotetun datan uudelleen hyödyntäminen vapaasti ja maksutta. Lähtökohtaisesti julkisen hallinnon keräämät tietoaineistot ovat julkisia, jollei laissa ole erikseen toisin säädetty. Julkisuusperiaatteen ja maksuttomuuden lisäksi datan avoimuutta määritteleviä tekijöitä ovat muun muassa aineiston tekninen saavutettavuus, löydettävyys ja uudelleen käytön sallivat lisensointiehdot. (Poikola ym. 2010, 11, 34, 47.)

3.1 Julkisuus

Julkisen hallinnon asiakirjojen ja muiden tietoaineistojen julkisuudesta säädetään julkisuuslaissa. Julkisuuslaki eli laki viranomaisten toiminnan julkisuudesta (621/1999; jäljempänä JulkL) säättää julkisuusperiaatteen lisäksi muun muassa asiakirjan saattamisesta yleisesti julkisuuteen, tiedonsaannista julkisesta asiakirjasta, tiedon antamisesta julkisesta asiakirjasta, salassapitovelvoitteista sekä viranomaisten velvollisuudesta edistää tiedonsaantia ja hyvää tiedonhallintatapaa. Avoimuus ja hyvä tiedonhallintatapa viranomaisten toiminnassa antaa yksilöille ja yhteisöille mahdollisuuden valvoa julkisen vallan ja julkisten varojen käyttöä, muodostaa vapaasti mielipiteensä sekä vaikuttaa julkisen vallan käyttöön ja valvoa oikeuksiaan ja etujaan (JulkL 1:3 §).

Julkisuusperiaatteen mukaan viranomaisten asiakirjat ovat julkisia, jollei laissa erikseen toisin säädetä (JulkL 1:1.1 §). Tässä laissa asiakirjalla tarkoitetaan ”kirjallisen ja kuvallisen esityksen lisäksi sellaista käyttönsä vuoksi yhteen kuuluviksi tarkoitetuista merkeistä muodostuvaa tiettyä kohdetta tai asiaa koskevaa viestiä, joka on saatavissa selville vain automaattisen tietojenkäsittelyn tai äänen- ja kuvantoistolaitteiden taikka muiden apuvälineiden avulla” (JulkL 1:5.1 §).

Asiakirja tai tietoaineisto on viranomaisen asiakirja, kun se on viranomaisen hallussa ja laadittu viranomaisen tai sen palveluksessa olevan toimesta tai viranomaisen antaman toimeksiannon johdosta (JulkL 1:5.2 §). Viranomaisen asiakirjana ei pidetä muun muassa viranomaisen palveluksessa olevan tai luottamushenkilön muun tehtävänsä tai asemansa vuoksi saamia kirjeitä tai asiakirjoja, muistiinpanoja tai luonnoksia, viranomaisen sisäistä käyttöä varten tarkoitettuja asiakirjoja tai löytötavaroiksi rinnastettavia asiakirjoja (JulkL 1:5.3 §).

Julkisuuslain piiriin kuuluvia viranomaisia ovat valtion hallintoviranomaiset, valtion ja eduskunnan virastot ja laitokset, tuomioistuimet ja muut lainkäyttöelimet, kunnalliset viranomaiset, valtion liikelaitokset sekä Suomen pankki, Rahoitustarkastus, Kansaneläkelaitos ja muut itsenäiset julkisoikeudelliset laitokset. Julkisuuden vaatimus koskee myös julkista tehtävää hoitavia yhteisöjä, laitoksia, säätiöitä ja yksityisiä henkilöitä niiden käyttäessä julkista valtaa. (JulkL 1:4 §.)

Oikeutta saada tieto viranomaisten julkisista asiakirjoista on rajoitettu yleisten ja yksityisten etujen suojaamiseksi (JulkL 1:2 §). Julkisuuslain 6 luvussa on erikseen säädetty yleistä asiakirjojen julkisuusperiaatetta rajoittavista salassapitoperusteista, vaitiolovelvollisuudesta ja hyväksikäyttökiellosta. Salassa pidettäviä asiakirjoja ovat esimerkiksi asiakirjat, joiden oikeudeton paljastuminen voi aiheuttaa vahinkoa maanpuolustukselle, valtion turvallisuudelle, kansainvälisille suhteille tai muille yleisille eduille (JulkL 8:36 §), tai joista tiedon antaminen olisi vastoin lapsen tai muuta erittäin tärkeätä yksityistä etua (JulkL 3:11.2 §). Asiakirjasalaisuutta koskevia säännöksiä sovellettaessa on otettava huomioon ettei tietojen saamista viranomaisen toiminnasta rajoiteta enempää kuin suojatavan edun vuoksi on tarpeellista ja ettei tiedon antamisesta ilmeisesti aiheudu haitallisia vaikutuksia (JulkL 5:17 §).

Julkisuuslain lisäksi tietoineiston julkisuuteen vaikuttaa tietosuojalainsäädäntö, jonka tarkoituksena on henkilötietojen suojaaminen oikeudettomalta ja henkilöä vahingoittavalta käytöltä (Poikola ym. 2010, 47). Henkilötietojen suojasta ja henkilötietojen käsittelystä säädetään henkilötietolaissa, tiettyjä aloja koskevissa erityissäännöksissä, laissa yksityisyyden suojasta työelämästä, sähköisen viestinnän tietosuojalaissa ja Euroopan unionin antamissa tietosuojaa koskevissa kansainvälisissä normeissa ja ohjeissa (Tietosuojavaltuutetun toimisto 2014, hakupäivä 8.4.2014).

Henkilötietolakia (523/1999; jäljempänä HenkL) sovelletaan henkilötietojen *automaattiseen käsittelyyn* ja myös muulloin silloin, kun henkilötiedot muodostavat henkilörekisterin tai sen osan (1:2.2 §). *Henkilötiedolla* tarkoitetaan luonnollista henkilöä koskevia merkintöjä, jotka voidaan tunnistaa häntä tai hänen perhettään koskeviksi, ja *henkilötietojen käsittelyllä* ”henkilötietojen keräämistä, tallettamista, järjestämistä, käyttöä, siirtämistä, luovuttamista, säilyttämistä, muuttamista, yhdistämistä, suojaamista, poistamista, tuhoamista sekä muita henkilötietoihin kohdistuvia toimenpiteitä” (HenkL 1:3.1 §). Lähtökohtaisesti henkilötietoja saa käsitellä ainoastaan rekisteröidyn antamalla suostumuksella tai rekisteröidyn toimeksiannosta tai kun käsittelystä säädetään laissa (HenkL 2:8.1 §).

Henkilötietolaki on yleislaki, johon nähden henkilötietojen käsittelyä koskevia erityissäännöksiä sovelletaan ensisijaisina. Erityislakeja tai -säännöksiä henkilötietolakiin nähden ovat esimerkiksi niin sanotut rekisterilait (väestötietolaki, laki terveydenhuollon valtakunnallisista henkilörekistereistä), viranomaisten tehtäviä ja toimintaa koskevat lait (laki verotustietojen julkisuudesta ja salassapidosta, laki potilaan asemasta ja oikeuksista), ja yksityisen sektorin rekisterinpitoon vaikuttavat lait (laki luottolaitostoiminnasta, vakuutustoimintaa koskevat lait). (Tietosuojavaltuutetun toimisto 2014, hakupäivä 8.4.2014.)

3.2 Maksuttomuus

Julkisin varoin tuotetun datan luovuttaminen uudelleen hyödynnettäväksi maksutta on merkittävä datan avoimuutta ja datan käyttöä lisäävä tekijä. Pienikin maksu nostaa kynnyksen tehdä kokeiluja data-aineistojen hyödyntämisestä ja näin rajoittaa uuden innovatiivisen toiminnan kehittelyä. Erityisesti tutkimus- ja koulutuskäytössä sekä aloittavien yritysten ja yksityisten aktiivien keskuudessa maksuttomuudella on merkitystä, koska kokeiluvaiheessa ei ole vielä varmuutta siitä, onko aineistolle pidemmällä aikavälillä käyttöä. Lisäksi tutkimus- ja koulutusorganisaatioissa usein tarvitaan valtuutus pientenkin hankintojen tekemiseen. (Poikola ym. 2010, 53, 57–58.)

Julkiselle sektorille data-aineistojen kerääminen ja ylläpito ei ole ilmaista, mutta tapahtuu verovaroin ja näin tuotettua dataa voidaan ajatella julkishyödykkeenä, jonka tulisi olla maksutta ja tasavertaisesti kaikkien saatavilla. Myös toimivan jakeluinfrastruktuurin rakentaminen maksaa ja datan jakelusta aiheutuneet lisäkustannukset onkin yksi peruste, miksi tietoluovutuksista on haluttu periä korvaus. (Poikola ym. 2010, 55, 59.)

Suomessa julkisen datan maksullisuuteen ja maksujen suuruuteen vaikuttaa vahvasti valtion maksuperustelaki, joka antaa yleiset perusteet valtion viranomaisten suoritteiden, kuten tavaroiden, palvelujen ja tietoluovutusten maksullisuudelle (Poikola ym. 2010, 48). Valtion maksuperustelaisissa (150/1992) valtion viranomaisten suoritteet jaetaan maksullisiin ja maksuttomiin suoritteisiin, julkisoikeudellisiin suoritteisiin sekä muihin suoritteisiin. Julkisoikeudellisella suoritteella tarkoitetaan ”valtion viranomaisen suoritetta, jonka kysyntä perustuu lakiin tai asetukseen ja jonka tuottamiseen viranomaisella on tosiasiallinen yksinoikeus” (1:3.2 §) ja josta perittävän maksun suuruuden tulee vastata suoritteen tuottamisesta aiheutuvaa omakustannusarvoa (2:6.1 §). 6 §:n 3 momentissa

mahdollistetaan kuitenkin maksun periminen omakustannusarvoa alempana tai kokonaan perimättä jättäminen perustelluista syistä.

Maksuperustelaki on asetettu vuonna 1992 eli ennen internetin ja digitaalisen tiedonvälityksen yleistymistä ja arkipäiväistymistä. Lain asettamisen aikaan tietoluovutuksille haluttiin asettaa hinta, joka kattaa sen tuottamisesta – tiedon etsimisestä, jäljentämisestä ja jakelusta – aiheutuvat kustannukset. Lain säätämisen taustalla vaikutti myös 1990-luvun alun lama ja sen aiheuttamat kustannustehokkuuspaineet julkiselle hallinnolle. Internetin mukanaan tuoman informaation jakelurakenteen muutoksen myötä tietoluovutusten kustannusperuste on muuttunut ja datan tarjoamisen kustannukset oleellisesti vähentyneet. (Poikola ym. 2010, 6, 56.)

Avoimen datan ajattelussa maksuperustaisuus rajoittaa julkisten digitaalisten aineistojen saatavuutta ja uudelleen käyttöä. Julkisen datan voidaan ajatella olevan samanlainen investointi kuin muukin julkisen hallinnon infrastruktuuri. Tällöin datan kerääminen ja julkaiseminen kustannettaisiin budjettivaroista ja raakadata jaettaisiin maksutta julkishyödykkeenä. (Poikola ym. 2010, 53, 59, 61.)

3.3 Koneluettavuus

Yksi avoimen datan peruskriteeri on aineiston tekninen saavutettavuus eli koneluettavuus. Eri verkkopalveluissa on tarjolla runsaasti tietoa, mutta suurin osa sellaisessa muodossa, ettei aineistoa voida helposti ohjelmallisesti jatkojalostaa. Esimerkiksi pdf-dokumenttien tai html-sivujen tietosisältöä koneet eivät pysty automatisoidusti käsittelemään ja eri lähteiden tietojen yhdistely on käytännössä mahdotonta. (Poikola ym. 2010, 18, 37.)

Datan julkaiseminen koneluettavassa muodossa mahdollistaa eri aineistojen yhteentoimivuuden ja siten laajojen ja monimutkaistenkin sovellusten kehittämisen (Dietrich ym. 2012, 6). Tyypillisiä esimerkkejä tällaisista monesta eri lähteestä dataa keräävistä ja koostavista palveluista ovat Mashup eli yhdistelmäpalvelut ja sijaintitietoja muihin avoimiin rajapintoihin yhdistävät palvelut. Kun tietojen käyttöön saamista helpotetaan teknisillä toimenpiteillä, julkista dataa voidaan hyödyntää huomattavasti laajemmin ja monipuolisemmin. Sen lisäksi, että data on saatavilla ohjelmallisesti käsiteltävissä muodossa, tiedon käyttäjille tärkeää on, että esitysmuoto on avoin eli aineistoa pystytään käyttämään mahdollisimman monilla käyttöjärjestelmillä ilman maksullisia ohjelmia. (Poikola ym. 2010, 18, 22, 66, 70.)

Yksinkertaisimmillaan dataa voidaan julkaista pilkku- tai sarkaineroteltuina tekstitiedostoina, jolloin dataa voidaan käsitellä taulukkolaskentaohjelmilla tai lukea koneellisesti. Esimerkiksi CSV-tiedostoformaattia (Comma Separated Value) voidaan hyödyntää taulukkomuotoisen datan tallentamisessa ja datan siirtämisessä toiseen tietokoneohjelmaan. (Poikola ym. 2010, 66.) Datan esittämiseen koneluettavassa muodossa liittyviä yleisen tason standardeja on selostettu taulukossa 1.

TAULUKKO 1. Standardeja datan esittämiseen ohjelmallisesti käsiteltävässä muodossa (Poikola ym. 2010, 64.)

XML	Extensible Markup Language (XML) on yleiskäyttöinen merkintäkieli, joka on laajennettavissa eri tarkoituksia varten uusilla merkintäelementeillä.
CSV	Comma Separated Value (CSV) tiedostomuoto, jossa arvot erotetaan toisistaan pilkuilla. Tiedostot voidaan avata taulukkolaskentaohjelmilla.
JSON	JavaScript Object Notation (JSON) on kevyt ohjelmointikielistä riippumaton tekstipohjainen datan siirtoformaatti.
RDF	Resource Description Framework (RDF) on linked data -paradigman standardi, jossa yksittäisiä tietoresursseja kuvaillaan niihin linkitettävien sanastojen avulla.

Koneluettavan muodon ohella on tärkeää, että data on riittävässä määrin rakenteinen eli strukturoitu ja tarjolla on dokumentaatio siitä, mitä rakenne pitää sisällään (Poikola ym. 2010, 37, 67). Rakenteinen tiedosto on tietokokonaisuus, jolle on ennalta määritelty rakenneosat eli elementit. Sallittu tietorakenne eli tietojen esittämisen järjestys voidaan määritellä rakennekuvauksena eli skeemana, joka on määrämuotoinen ja tietokoneen ymmärrettävissä oleva muoto. Rakenteeseen voidaan liittää myös esitystapaan, julkaisukanavaan ja käyttöoikeuksiin liittyviä sääntöjä ja automatisoituja toimintoja. (Kaario ym. 2008, 39.)

3.4 Löydettävyys

Datan julkaisussa oleellista on tehdä siitä mahdollisimman löydettävää. Jos aineiston olemassaolo ei ole datavarantoja käyttävien tahojen tiedossa, tavoitellut hyödyt jäävät saavuttamatta. Koneluettavuuden lisäksi data-aineiston löydettävyyttä voidaan tehostaa käyttämällä hyvin järjestettyjä ja semanttisesti yhteentoimivia metatietoja ja lisäämällä data julkiseen datakatalogiin. Myös perinteinen tiedottaminen ja markkinointi verkossa, julkaisuissa ja tapahtumien yhteydessä on erinomainen keino kertoa datan olemassaolosta. (Poikola ym. 2010, 27–28, 35.)

Metatiedoilla on olennainen merkitys data-aineistojen löytymiselle ja käsittelylle. Sen lisäksi että hakukoneet pystyvät etsimään tietoa tarkemmin ja monipuolisemmin, yhtenäiset metatiedot mahdollistavat tietojärjestelmien väliset tiedonsiirrot ja eri lähteissä olevien sisältöjen yhdistämisen. Metatieto eli kuvailutieto on rakenteista tietoa tiedosta ja teknisesti erotettavissa varsinaisesta tietosisällöstä. Semanttisen ja teknisen yhdenmukaisuuden varmistamiseksi on tärkeää, että metatiedoissa noudetaan alan standardeja, suosituksia ja sanastoja. (Kaario ym. 2008, 25–27.)

Datakatalogi on strukturoitu metatietorekisteri, johon on yhdistetty metatietoa useampien julkisten organisaatioiden hallussa olevista aineistoista. Datakatalogit voivat olla kansallisia, seudullisia, kaupunkien ylläpitämiä tai yksityisten tahojen ylläpitämiä. Kun lista organisaation datavarannoista on julkaistu, voidaan siitä viestiä potentiaalisille datan uudelleenkäyttäjille, luoda kontakteja ja kerätä mielipiteitä. Läsnäolo sosiaalisessa mediassa on luonteva tapa tavoittaa kehittäjäyhteisöjä ja datan hyödyntämisestä kiinnostuneita tahoja. (Poikola ym. 2010, 40, 75.)

3.5 Uudelleenkäytettävyys

Avoimen datan vapaaseen uudelleen hyödyntämiseen liittyy keskeisesti aineistoihin mahdollisesti kohdistuvat tekijänoikeudet ja lisensointikäytännöt. Avointa dataa julkaistaessa aineiston käytölle, muokkaukselle ja uudelleenjakelulle ei tule asettaa rajoittavia käyttö- tai lisensointiehtoja ja aineistoon mahdollisesti kohdistuvista tekijänoikeuksista luovutaan. Datan hyödyntäjille annetaan lailliset oikeudet datan uudelleenkäyttöön sisällyttämällä aineiston yhteyteen avoin käyttö lupa (lisenssi), josta ilmenee aineiston käyttöehdot ja alkuperä. (Poikola ym. 2010, 36, 45.)

Elinkeinoelämän tutkimuslaitoksen (ETLA) julkaisemassa tutkimuksessa (Salokannel 2012, 2) on tarkasteltu julkisen datan tekijänoikeuksia ja lisensoinnin lähtökohtia. Julkisen datan uudelleen hyödyntämistä hidastaa Suomen laaja tekijänoikeussuoja, joka aiheuttaa epäselvyyttä ja oikeudellista epävarmuutta käyttöoikeuksista. Vaikka laajasta tekijänoikeussuojasta aiheutuva oikeudellinen epävarmuus on poistettavissa selkeän avoimen lisenssin avulla, julkisen sektorin viranomaisella on oltava kaikki oikeudet julkaistavaan dataan, jotta aineistolle voidaan myöntää avoin käyttö lupa. Eli jos aineistoon sisältyy kolmansien tahojen tekijänoikeudellista materiaalia, julkisen viranomaisen on hankittava oikeudet myös tämän materiaalin lisensoimiseen.

Julkisin varoin tuotettu tieto voidaan karkeasti jaotella tekijänoikeudellisesti suojattuihin yksittäisiin teoksiin sekä erilaisiin tietokannoista tai luetteloina saatavilla oleviin tietoihin. Teossuojan piirissä olevaa materiaalia ovat esimerkiksi työryhmien raportit tai vastaavat kirjalliset teokset. Tietokannoissa oleva raakadata sellaisenaan on tekijänoikeudellisen suojan ulkopuolella, mutta raakadataa sisältävät tietokannat ja luettelot voivat tulla suojatuiksi joko tekijänoikeuslaissa annettujen erityissäännöksiä perusteella (tekijänoikeuden lähioikeudet muun muassa luetteloille, taulukoille ja tietokannoille, joissa on yhdisteltynä suuri määrä tietoja) tai teostason ylittyessä kirjallisena teoksena. Pääsääntöisesti julkisilla tietoaineistoilla ja raakadatalalla on ainakin tekijänoikeuden lähioikeuden mukainen suoja. Ennen julkisen datan lisensointia on kuitenkin arvioitava onko aineisto tekijänoikeudellisesti suojattu ja onko julkaisijalla kaikki oikeudet julkaistavaan aineistoon. (Salokannel 2012, 15, 18, 21.)

Aineiston käyttölisenssin tarkoituksena on antaa käyttäjälle oikeudellinen varmuus siitä, miten lisenssin kohteena olevaa dataa voi käyttää syyllistymättä tekijänoikeuksien loukkaukseen. Aineiston vapaan uudelleenkäytön sallimiseen voidaan käyttää avoimia lisenssejä (esimerkiksi Creative Commons tai Open Data Commons), jotka asettavat aineiston käytölle mahdollisimman vähän rajoituksia tai joiden avulla voidaan luopua kaikista aineistoon automaattisesti kohdistuvista oikeuksista. (Poikola ym. 2010, 49, 89.) Esimerkiksi Julkisen hallinnon tietohallinnon neuvottelukunnan (JUHTA 2014, 2, 13.) valmistelemissa avoimen tietoaineiston käyttöluvassa julkishallinnon avoimiin tietoaineistoihin suositellaan Creative Commons Nimeä 4.0 –lisenssiä (CC BY 4.0), joka sallii aineiston vapaan käytön kaikin mahdollisin tavoin edellyttäen, että aineiston lähde nimetään käyttöluvan antajan edellyttämällä tavalla. Silloin kun aineiston yksilöiminen ja tuottajan tietojen ilmoittaminen ei ole tarpeellista, suositellaan käytettäväksi Creative Commons Zero –lisenssiä (CC0), jolloin aineisto on vapaasti käytettävissä ja jalostettavissa ilman mitään rajoitteita.

Julkaistavan datan lisäksi myös aineistojen käyttölisenssit suositellaan julkaistavaksi koneluettavassa muodossa. Lisenssien koneluettavuus helpottaa eri lähteistä yhdisteltävien aineistojen käyttöehtojen huomioimista ja noudattamista sekä parantaa lisensoitujen aineistojen löydettävyyttä. (Poikola ym. 2010, 67.) Esimerkiksi Creative Commons -lisenssit voidaan liittää digitaalisen aineiston metatietoihin xmp-muodossa tai html-sivulle koneluettavana RDFa-merkittynä html-koodina (Creative Commons Suomi 2014, hakupäivä 23.9.2014).

4 DATAN JULKAISUPROSESSI

Datan julkaisu kaikille avoimeksi dataksi alkaa omien aineistojen analysoinnilla. Ensimmäinen lista organisaation datavarannoista sisältää olennaisimmat tiedot organisaation hallussa olevista aineistoista, kuten aineiston nimi ja sisällön kuvaus sekä nykytilanteen arviointi eli kuinka hyvin aineisto täyttää avoimuuden kriteerit tällä hetkellä. Ennen julkaisua on selvitettävä aineistoihin liittyvät omistajuuskysymykset ja varmistettava, ettei julkaiseminen ole julkisuus-, henkilötieto- tai tekijänoikeuslain vastaista. Datavarannot voidaan jakaa täysin julkisiin, epäselviin ja ei-julkisiin aineistoihin ja aloittaa datan avaaminen selvästi julkisista aineistoista. Kun omat datavarannot on tunnistettu, niistä voidaan valita avattaviksi ensin sellaiset, jotka ovat jo valmiiksi helposti julkaistavassa muodossa tai joilla tiedetään jo valmiiksi olevan kysyntää. Ensimmäinen julkaistava data voi myös olla pelkkä listaus organisaation julkisista datavarannoista, jotka ovat avattavissa tai joiden avoimuutta voidaan jatkossa lisätä. Kun listaus datavarannoista on julkaistu avoimesti verkossa, voidaan siitä viestiä potentiaalisille datan uudelleenkäyttäjille ja kerätä mielipiteitä siitä, mille datalle on eniten kysyntää. (Poikola ym. 2010, 39–41.)

Datavarantojen kartoittamisen jälkeen arvioidaan minkälaisin käyttöehdoin ja missä muodossa avattavaksi valittu data halutaan julkaista (Dietrich ym. 2012, 7). Oikeudellisten epäselvyyksien välttämiseksi avoin lisenssi on hyvä antaa myös aineistoille, jotka ovat tekijänoikeussuojan ulkopuolella. Lisenssi suositellaan liitettäväksi aineistoon sekä ihmislueuttavassa muodossa verkkosivulla, jossa aineisto julkaistaan, että kone-lueuttavana metadatanä. (Toikkanen 2014, hakupäivä 3.10.2014.) Pilotointivaiheessa eli ensimmäisissä datan avauksissa riittää, kun raakadata tarjotaan pelkistetysti pilkku- tai sarkaineroteltuina tekstitiedostoina tai valmiina taulukkolaskentatiedostoina. Jos kiinnostus dataa kohtaan kasvaa, datan käytettävyyttä voidaan lisätä tarjoamalla aineisto myös XML-, JSON- tai RDF-muodossa, jotka soveltuvat hyvin monimutkaisempien tietorakenteiden esittämiseen ja siirtämiseen. (Poikola ym. 2010, 42, 66–67.)

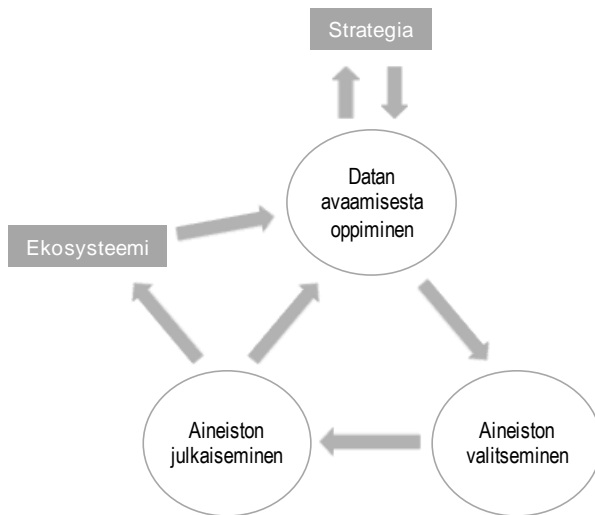
Seuraavassa vaiheessa tehdään datasta helposti saavutettavaa ja löydettävää. Yksinkertaisin datan jakelutapa on ladattava tiedosto organisaation omilla verkkosivuilla. Datan kysynnän kasvaessa aineiston jakelua voidaan kehittää toteuttamalla avoin ohjelmointi- tai palvelurajapinta (API, Application Programming Interface), jonka kautta voidaan ohjelmallisesti hakea raakadataa. Tyypillisesti rajapinnan kautta ollaan yhteydessä tietokantaan, jolloin etuna ovat tietojen ajantasaisuus sekä käyttäjän mahdollisuus valita käyttöönsä vain haluamansa osa datasta. Datan julkaisussa on myös

mahdollista hyödyntää datan tallennuspalveluja, jotka ovat yleensä tietyn yhteisön tarpeisiin muodostuneita. Tallennuspalvelujen etuna ovat valmiit aiheesta kiinnostuneet verkostot, mutta on mahdollista, että pääsyä datasisältöön on rajoitettu, jolloin avoimen datan kriteerit eivät täyty. (Dietrich ym. 2012, 9–10.) Avoimuuden kriteerit täyttävässä datan luovutuksessa datan pitää olla kokonaisuudessaan vapaasti ladattavissa internetin kautta, eikä datan käyttömahdollisuuksia ole rajoitettu käyttäjien tai käyttötarkoitusten mukaan. Esimerkiksi datan tarjoaminen vain tutkimus- ja tuotekehityskäyttöön tai ainoastaan epäkaupalliseen käyttöön asettaa käyttäjät eriarvoiseen asemaan, eli tällöin ei ole kyse avoimesta datasta sen varsinaisessa merkityksessä. (Poikola ym. 2010, 35.)

Datan avaamisesta on hyötyä vasta kun dataa käytetään. Datan avaaminen ei ole päämäärä itsessään, vaan keino synnyttää uusia luovia datan käyttötapoja ja liiketoimintamahdollisuuksia. Datan avaamista voi ajatella vuorovaikutteisena prosessina, jossa palvelua suunnitellaan käyttäjälähtöisesti. Datan käyttäjillä on paras tuntemus siitä, kuinka dataa kannattaa tarjota, jotta sitä aktiivisesti hyödynnettäisiin. Siksi on tärkeää huomioida jo mahdollisimman varhaisessa vaiheessa, miten potentiaaliset avoimen datan hyödyntäjät tavoitetaan. Avattuja datavarantoja listaavista datakatalogeista on tullut hyvin suosittu ja yleinen keino luoda näkyvyyttä avatulle datalle, mutta myös suorat kontaktit kehittäjäyhteisöihin sosiaalisessa mediassa sekä erilaisissa tapahtumissa ja kilpailuissa ovat hyvä kanava lisätä datavarantojen tunnettuutta. (Poikola ym. 2010, 21, 25; Dietrich ym. 2012, 10–13.)

Iteratiivinen prosessi

Datan avaaminen vaatii aina aikaa ja resursseja, ja avaamisessa kannattaakin lähteä liikkeelle pienin askelin. Kaikkea ei tarvitse tehdä kerralla valmiiksi, vaan avaaminen voidaan aloittaa pienistä ja yksinkertaisista aineistoista edeten vaiheittain kohti laajempia kokonaisuuksia. Vaihe vaiheelta kerrytetään käytännön kokemuksia avoimeen dataan liittyvästä toiminnasta ja ekosysteemiin osallistumisesta (kuvio 3). Ohjelmakehityksessä yleistyneet ketterät menetelmät soveltuvat hyvin lähitökohtaksi datan avaamiseenkin. Vuorovaikutuksessa avoimen datan ekosysteemin kanssa datan avaamista kehitetään iteratiivisesti käyttökokemuksiin ja palautteisiin reagoiden. (Poikola ym. 2010, 26, 38.) Avoimen datan ekosysteemi käsittää kaikki avoimen datan keräämiseen, jalostukseen, julkaisuun ja uudelleenkäyttöön osallistuvat tahot ja järjestelmät, jotka ovat keskenään vuorovaikutuksessa ja muokkaavat ekosysteemin käytäntöjä (Poikola ym. 2010, 15).



KUVIO 3. Datan avaamista voidaan kehittää iteratiivisesti (Poikola ym. 2010, 40)

Datavarantojen listauksesta alkaneessa prosessissa pyritään uusien kierrosten aikana parantamaan datan sisällön ja käyttöoikeuksien kuvailua ja dokumentaatiota sekä kehittämään datan jakelua. Datavarantojen yhteiskäyttöisyyttä ja löydettävyyttä lisätään metatietosuunnittelun ja linkitettävien kuvailusanastojen avulla. Analysoimalla käytössä olevia datavarantoja syvennetään tuntemusta, missä ja miten organisaatiossa dataa tuotetaan, ja inventaarion tuloksena havaittua hajallaan olevaa tai päällekkäistä dataa voidaan harmonisoida. Datan julkaisusta saadun kokemuksen karttuessa organisaatiolle voidaan muodostaa datan avoimuutta edistävä strategia, jonka mukaisesti datavarantoja ja rajapintoja järjestelmällisesti avataan. Ajan myötä suunnitelmallinen datavarantojen kartoittaminen ja datan avaaminen on osa organisaation tietojärjestelmien ja tietoarkkitehtuurin kehittämistyötä. (Poikola ym. 2010, 42–43.)

5 DATAN AVAUKSIA SUOMESSA

Suomessa julkinen hallinto on ryhtynyt aktiivisesti edistämään avoimen julkisen datan hyödyntämistä, saatavuutta ja käyttöä. Valtiovarainministeriön asettaman ja organisoiman avoimen tiedon ohjelman tavoitteena on purkaa julkisen datan uudelleenkäyttöä rajoittavia esteitä, luoda edellytyksiä tietovarantojen suunnitelmalliselle avaamiselle ja kehittää yhteistyötä eri avoimen tiedon toimijoiden kesken. Osana avoimen tiedon ohjelmaa julkaistiin syksyllä 2014 uusi kansallinen datakatalogi avoindata.fi, joka korvaa aikaisemman data.suomi.fi-katalogin. Avoindata.fi-portaalista löytyy yhden osoitteen alta julkisen hallinnon avaamaa dataa sekä yhteentoimivuutta edistäviä ohjeita ja työkaluja. (Valtiovarainministeriö 2014, hakupäivä 4.12.2014.)

Datan avaamiseen on kannustettu myös paikallistasolla tiedon avoimuutta ja avoimien aineistojen käyttöä edistävän Open Knowledge Finland -yhdistyksen organisoimilla Avoimen tiedon kuntakieretuilla. Näitä Open Knowledge Roadshow -tapahtumia on järjestetty syksyllä 2013 ja 2014 Jyväskylässä, Kuopiossa, Mikkelissä, Oulussa, Porissa, Rovaniemellä, Tampereella ja Turussa. Kiertueen tavoitteena on käynnistää ja tukea paikallista toimintaa sekä vahvistaa avoimen tiedon verkostoa maanlaajuisesti. (OKFN LOCAL: Finland 2014, hakupäivä 5.12.2014.) Syksyn 2013 kuntakierteen yhteydessä ja innoittamana esimerkiksi Oulun kaupunki julkisti oman Oulu open data -sivuston (Kaleva 2013, hakupäivä 5.12.2014; Open Knowledge Finland 2013, hakupäivä 5.12.2014).

5.1 Avoindata.fi-portaali

Avoindata.fi-palvelun kautta on löydettävissä yli 1200 avoin data -tietoaineistoa (hakupäivä 4.12.2014). Tietoaineistoja voi etsiä sekä aiheittain, sisältötyypeittäin että tiedon tuottajan tai tiedostomuodon mukaan. Tietoaineistoista suurimman kokonaisuuden muodostaa Helsinki Region Infoshare -palvelun ylläpitämän datakatalogin data. Datakatalogi sisältää huomattavan määrän data-aineistoa Helsingin seudusta, asukkaista ja palveluista: muun muassa tilasto- ja ennustetietoa, paikkatietoa ja erilaisia karttoja. Toinen avoindata.fi-palvelun suurempi kokonaisuus muodostuu valtionhallinnon avoimista paikkatietoaineistoista. Paikkatietohakemiston kautta löytyy muun muassa digitaalisia maastotieto-, karttakuva- ja nimistöaineistoja, sekä Suomen tie- ja katuverkon sijainti- ja ominaisuustietoaineistoja. Taulukossa 2 havainnollistetaan avoindata.fi-palvelun tietoaineistoja joulukuussa 2014. Kaikki julkaistut tietoaineistot eivät ole vielä koneluettavassa muodossa, vaan ovat tarjolla esimerkiksi html- tai pdf-dokumentteina avoimella lisenssillä.

TAULUKKO 2. Avoindata.fi-palvelun avoin data -tietoaineistot joulukuussa 2014

Tiedon tuottaja	Tieto-aineistoja	Tietoaaineistojen kuvaus	Tiedosto-muoto
Ulkoinen lähde: Hri.fi Verkkopalvelun osoite:	1128	Pääkaupunkiseudun kaupunkien julkisia tietovarantoja avoimena datana http://www.hri.fi/	rajapinta-palvelu, useita tiedostomuotoja
Ulkoinen lähde: Paikkatietohakemisto Verkkopalvelun osoite:	61	Valtionhallinnon avoimia paikkatietoaaineistoja http://www.paikkatietohakemisto.fi http://www.paikkatietoikkuna.fi/web/fi/avoin-paikkatieto	rajapinta-palvelu, html, pdf
Tampere	11	Tampereen kaupungin avoimet tietoaaineistot	rajapinta-palvelu, useita tiedostomuotoja
Oikeusministeriö	2	Hallinto- ja hovioikeuksien ratkaisuja	html
Valtiokonttori	2	Valtion budjettitalous, Valtion henkilöstö	csv
Valtiovarainministeriö	2	Valtion avattuja tietovarantoja 2014, Valtion talousarvioesitykset	csv, xml
Maa- ja metsätalousministeriön tietopalvelukeskus	1	Maatilatilastollinen vuosikirja 2013	excel
Suomi syö ja juo -hanke	1	Valokuvia suomalaisesta ruoka- ja juomakulttuurista eri aikakausilta	jpg, png, tiff
Valtioneuvoston kanslia	1	Hallitusohjelman seurantadata	csv, pdf
Väestötietokeskus	1	Väestötietojärjestelmän suomalaisten nimiaineistot	csv, pdf

5.2 Muita avoimen datan lähteitä

Julkisen hallinnon virallisen datakatalogin rinnalla toimii lukuisia paikallisia ja organisaatiokohtaisia sivustoja. Jo aiemmin mainitun Oulun kaupungin avoin data -sivuston lisäksi omia sivustojaan ovat perustaneet esimerkiksi Jyväskylä, Mikkeli ja Tampere. Turun ja ympäristöalueiden avointa dataa kerää alueellinen paikkatietoyhteistyöverkosto Lounaispaikka. Myös useat valtionhallinnon organisaatiot julkaisevat avoimia aineistojaan omilla sivustoillaan ilman että ovat linkittäneet niitä viralliseen avoindata.fi-portaaliin. Taulukkoon 3 on koostettu eri toimijoiden avoin data -sivustoja ja niiden sisältöä.

TAULUKKO 3. Paikallisia ja organisaatiokohtaisia avoin data -sivustoja joulukuussa 2014

Tiedon tuottaja	Tietoaineistojen kuvaus	Tiedosto- muoto
Ilmatieteen laitos Verkkopalvelun osoite: https://ilmatieteenlaitos.fi/avoin-data	<ul style="list-style-type: none"> - sää-, meri- ja ilmastohavaintoja - säätutkakuja ja salamahavaintoja - kansallisen sääennuste- ja merimallien tietoja 	latauspalvelu, rajapintapalvelu
Liikennevirasto Verkkopalvelun osoite: http://www.liikennevirasto.fi/avoindata	<ul style="list-style-type: none"> - kelikamerakuvia sekä ajantasaisia liikenteen sujuvuus- ja häiriötietoja maantieverkolta - paikkatietoaineistoja: Suomen tie- ja katuverkon aineistot, bussipysäkit, rataverkko ja tasoristeykset - merikartta-aineistoa - tietoja kotimaan juna-, lento- ja linja-autoliikenteestä 	latauspalvelu, rajapintapalvelu
Maanmittauslaitos Verkkopalvelun osoite: http://www.maanmittauslaitos.fi/avoindata	<ul style="list-style-type: none"> - maastotietoaineistoja: ortoilmakuvia, korkeusmalleja, vektori- ja rasterikarttoja 	latauspalvelu, rajapintapalvelu
Liikenteen turvallisuusvirasto Trafi Verkkopalvelun osoite: http://www.trafi.fi/tietopalvelut/avoin_data	<ul style="list-style-type: none"> - ajoneuvojen teknisiä tietoja sisältävä pilottiaineisto: liikennekäytössä olevien ajoneuvojen rekisteröinti-, hyväksyntä- ja teknisiä tietoja 	csv
Verohallinto Verkkopalvelun osoite: http://vero.fi/fi-FI/Avoin_data(29357)	<ul style="list-style-type: none"> - yhteisöjen tuloverotuksen julkiset tiedot 	csv
Jyväskylän kaupunki Verkkopalvelun osoite: http://data.jyvaskyla.fi/	<ul style="list-style-type: none"> - kaupungin päätöksentekoon ja palveluihin liittyviä aineistoja - joukkoliikenteen reitti- ja aikataulutietoja 	excel, txt, wms (paikkatieto) html, rss, word
Mikkelin kaupunki Verkkopalvelun osoite: http://open.mikkeli.fi/	<ul style="list-style-type: none"> - kaupungin päätöksentekoon ja palveluihin liittyviä aineistoja - paikkatietoaineistoja 	JSON-rajapinta, csv, excel jpg, pdf
Oulun kaupunki Verkkopalvelun osoite: http://www.ouka.fi/oulu/oulu-tietoa/avoin-data	<ul style="list-style-type: none"> - kaupungin päätöksentekoon ja palveluihin liittyviä aineistoja - joukkoliikenteen reitti- ja aikataulutietoja 	rajapinta, csv, excel, gfts, xml
Tampereen kaupunki Verkkopalvelun osoite: http://tampere.fi/tampereinfo/avoindata.html	<ul style="list-style-type: none"> - kaupungin päätöksentekoon ja palveluihin liittyviä aineistoja - kartta- ja paikkatietoaineistoja - joukkoliikenteen reaaliaikainen rajapinta 	rajapinta, csv, excel, json, tiff, wms pdf, rss, word
Lounais-Suomen paikkatietokeskus Verkkopalvelun osoite: http://paikkatietokeskus.lounaispaikka.fi/fi/aineistot/	<ul style="list-style-type: none"> - kartta- ja paikkatietoaineistoja - liikenneaineistoja - kaavoitus- ja rakennustietoja 	rajapinta, excel MapInfo TAB, GeoJSON
Aalto-yliopisto Verkkopalvelun osoite: http://data.aalto.fi/	<ul style="list-style-type: none"> - palvelu keskeneräinen 	

Metropolia Ammattikorkeakoulu Verkkopalvelun osoite:	– opetussuunnitelmat ja niihin liittyvät tiedot – opintojaksosten toteutukset ja niihin liittyvät tiedot – varaukset eli lukujärjestystiedot – tilojen tiedot http://opendata.metropolia.fi/index.html https://wiki.metropolia.fi/display/opendata/Open+data+-+Metropolia+Ammattikorkeakoulussa	rajapintapalvelu
HelMet-kirjastot Vaski-kirjastot Verkkopalvelun osoite:	– kirjastojen aineistotietokanta Info:	MARCXML
	http://data.kirjastot.fi/data.html http://www.hri.fi/fi/dataset/helsingin-kaupunginkirjasto-avoin-data-aineistoluettelot-helmet-kirjastojen-luettelotiedot https://www.vaskikirjastot.fi/web/arena/avoin-data	

5.3 Esimerkkejä avoimen datan hyödyntämisestä

Useimmat avointa dataa julkaisevat organisaatiot ovat kiinnostuneita, kuinka heidän julkaisemaansa dataa hyödynnetään, ja kokoavat sivuilleen esimerkkejä ja ideoita avoimien aineistojensa käytöstä. Myös Apps4Finland-kilpailun kotisivut, <http://www.apps4finland.fi/>, ovat erinomainen kanava tutustua erilaisiin tapoihin hyödyntää avointa dataa.

Mashup-tyyppisiä yhdistelmäpalvelusovelluksia

Carpark Oulu (<http://iiska.github.io/harbour-carpark-oulu/>) on mobiilisovellus Jolla-puhelimelle, joka näyttää vapaiden pysäköintipaikkojen lukumäärät Oulussa sijaitsevissa parkkihalleissa. Palvelu hyödyntää Oulunliikenne.fi -palvelun tarjoamaa pysäköintitilojen ajantasaisia käyttötietoja.

HelMet Viivakoodinlukija (<https://play.google.com/store/apps/details?id=org.apps4fi.HelMet>) on Android-laitteille tehty sovellus, joka hyödyntää HelMet-kirjastojen aineistotietokantoja sekä pääkaupunkiseudun kirjastojen sijaintitietoja. Viivakoodinlukijalla voi etsiä kirjan viivakoodin avulla lähimmän kirjaston, josta kirja on lainattavissa.

Tilannehuone.fi on kotimainen uutispalvelu, joka hyödyntää avointa dataa välittäessään reaaliaikaista tietoa pelastustoimen tehtävistä Suomessa (kuvio 4). Sovellus muun muassa yhdistää hälytyskeskusten tilannetiedotukset ja paikkatiedot Google-kartalla. Palvelu hyödyntää myös Liikenneviraston liikennetiedotteita.

Location	Time	Description
Kotka	07.12.2014 19:55:28	palohälytys
Kerava	07.12.2014 19:43:30	tulipalo muu: pieni
Helsinki	07.12.2014 19:24:39	eläintehtävät: eläimen pelastaminen
Kuopio	07.12.2014 19:21:34	tieliikenneonnettomuus: pieni
Lahti	07.12.2014 19:19:09	liikennevälinepalo: pieni
Siikajoki	07.12.2014 19:14:30	tieliikenneonnettomuus: pieni
Oulu	07.12.2014 18:59:46	rakennuspalo: pieni

KUVIO 4. Tilannehuone.fi-palvelun näkymä

Datan visualisointeja

Potentiaaliset menetetyt elinvuodet -visualisaattori (<http://open.mikkeli.fi/pyllldata/>) visualisoi ennenaikaisesti menetettyjä elinvuosia. Palvelu käyttää Mikkelin kaupungin julkaisemaa data-aineistoa, jossa henkilön ikää kuolinhetkellä verrataan hänen laskennalliseen elinajanodotteeseen.

Rakennettu Helsinki 1812–2012 (<http://www.hri.fi/fi/sovellukset/rakennettu-helsinki-1812-2012/>) visualisoi Helsingin kaupungin kehitystä kartalla. Sovelluksella voi tarkastella tietoja rakennusten valmistumisvuosista, tarkastella kaupunkinäkömää vuosiluvun mukaan tai seurata kaupungin kasvua animaationa. Palvelu hyödyntää Helsingin kaupungin rakennusrekisteristä koostettua dataa. Sovellus palkittiin vuonna 2013 Apps4Finland-kilpailun Hahmota-sarjan parhaana työnä.

6 JOHTOPÄÄTÖKSET

Tutkimustehtävän tarkoitus oli perehtyä uuteen, vähän tunnettuun ilmiöön ja selvittää siihen liittyviä käsitteitä ja määritelmiä. Tutkimuksessa haettiin vastauksia kysymyksiin, mitä tarkoitetaan avoimella datalla ja tietovarantojen avaamisella, sekä mitä vaatimuksia tietovarantojen avaamisessa on otettava huomioon. Lisäksi haluttiin kartoittaa, millaisia tietovarantoja Suomessa on avattu. Tutkimuksen tietoperusta rakennettiin pohjautuen aiheeseen liittyviin julkaisuihin, artikkeleihin, muistioihin ja raportteihin. Tutkimusaiheen tuoreudesta johtuen teoreettista kirjallisuutta tai aikaisempia tutkimusjulkaisuja ei ollut käytettävissä.

Avoimen datan määritelmä pitää sisällään useita käsitteitä, kuten laillinen ja vapaa uudelleenkäytettävyys, maksuttomuus, koneluettavuus ja löydettävyys. Data-aineisto on avointa, kun kuka tahansa voi sitä vapaasti käyttää, muokata ja uudelleen jakaa. Datan avoimuutta määriteltäessä korostuu erityisesti koneluettavuus eli ohjelmallisesti käsiteltävä avoin esitysmuoto ja maksuttomuus. Tarjoamalla data koneluettavassa muodossa mahdollistetaan data-aineistojen laaja ja monipuolinen käyttö ja uusien digitaalisten palvelukanavien syntyminen. Aineistojen luovutuksesta perittävä pienikin maksu nostaa kynnyistä tehdä kokeiluja data-aineistojen hyödyntämisestä erityisesti tutkimus- ja koulutusikäikäytössä, ja näin rajoittaa uuden innovatiivisen toiminnan kehittelyä.

Tietovarantojen avaamisessa on huomioitava, että julkaistavat data-aineistot ovat lainsäädännön puitteissa julkisia, ja että aineiston hyödyntäjillä on laillinen oikeus aineiston uudelleenkäyttöön. Tietovarantojen julkisuutta määrittävät muun muassa julkisen hallinnon asiakirjojen ja muiden tietoineistojen julkisuudesta säätävä julkisuuslaki, henkilötietojen suojasta ja henkilötietojen käsittelystä säätävä henkilötietolaki, sekä henkilötietojen käsittelyä koskevat erityislait ja -säännökset. Lailliset oikeudet tekijänoikeussuojan alaisen datan uudelleenkäyttöön annetaan avoimella käyttöluvalla eli lisenssillä, mutta on huomioitava, että jos aineistoon sisältyy kolmannen tahon tekijänoikeudellista materiaalia, data-aineiston julkaisijan on hankittava oikeudet myös tämän materiaalin lisensoimiseen.

Suomessa julkisten tietovarantojen avaamisessa ollaan oltu edelläkävijöiden joukossa, ja avoimen datan edistämiseksi tehdään aktiivisesti töitä. Pääkaupunkiseudun kaupungit eli Helsinki, Espoo, Vantaa ja Kauniainen ovat avanneet huomattavan määrän kaupunkien päätöksentekoon ja palveluihin liittyviä data-aineistoja yhteisen avoimen datan palvelunsa kautta, ja ovat toiminnallaan olleet

esimerkkinä useille muille kaupungeille. Julkisen hallinnon organisaatiot ovat avanneet vuodesta 2012 lähtien muun muassa maasto- ja paikkatietoaineistoja, sää-, ilmasto- ja meridataa sekä liikenne- ja viestintädataa. Avoimelle datalle myönteisestä suuntauksesta huolimatta korkeakoulut eivät ole vielä aktivoituneet tietovarantojen avaamisessa.

Tietovarantojen avaamisella on edistetty arkea helpottavien ja hallinnon läpinäkyvyyttä lisäävien palveluiden syntymistä. Avointa dataa on hyödynnetty erilaisissa verkko- ja mobiilisovelluksissa sekä visualisoinneissa ja analyyseissä. Suosittuja aineistoja ovat olleet paikkatieto- ja liikenneaineistot, joiden pohjalta on kehitetty esimerkiksi useita julkisen liikenteen käyttöä helpottavia mobiilisovelluksia. Kuinka paljon kokonaan uutta liiketoimintaa ja innovatiivista toimintaa avoimen datan ympärille kehittyi, nähdään vasta, kun tietovarantojen avaamisesta saadaan enemmän kokemusta ja tutkimustietoa.

7 POHDINTA

Tämän opinnäytetyön toteuttaminen on ollut mielenkiintoinen tutustumismatka tietojenkäsittelijöille ajankohtaiseen aiheeseen. Tutkimuksessa oli haasteellista aiheen tuoreus ja käytettävissä olleen lähdemateriaalin niukkuus ja pirstaleisuus. Tätä työtä voikin hyvin verrata palapelin kokoamiseen, jossa erillisistä paloista pyrittiin muodostamaan ehjä kokonaisuus. Oikeiden palojen eli lähdemateriaalin valinnassa on pitänyt olla kriittinen ja analyttinen, ja usein yksi yksittäinen palanen on johdattanut toisen luo. Tästä johtuen tämä tutkimusprosessi on ollut työläs ja ajallisesti vaativa. Mielestäni onnistuin kuitenkin kokoamaan yhteen aiheeseen liittyvät keskeiset käsitteet, määritelmät ja ilmiöt. Tutkimusprosessin seurauksena muodostui looginen ja selkeä kokonaisuus, joka vastaa asetettuihin tutkimuskysymyksiin.

Koska julkisten tietovarantojen avaamisella halutaan edistää muun muassa uuden liiketoiminnan syntymistä uusien palveluiden ja sovellusten muodossa, ja koska avoimen datan jalostamismenetelmät liittyvät keskeisesti tietojenkäsittelyyn, avoimen datan hyödyntäminen opetuksessa olisi perusteltua. Koulutusvaiheessa saatu perustieto avoimesta datasta ja käytännön harjoittelu avatuilla data-aineistoilla antaisi mahdollisesti kipinän luoda uusia innovatiivisia palveluja, joita voisi kaupallisesti hyödyntää.

Lähdemateriaalia kerätessä vastaan tuli useita avoimeen dataan läheisesti liittyviä ilmiöitä, jotka jätin tämän tutkielman ulkopuolelle pysyäkseeni omassa aiheen rajauksessani. Linkitetty data ja semanttinen verkko, pilvipalvelut, Internet of things, Big data, SOME-data, My data; nämä kaikki liittyvät tietolähteiden avautumiseen, digitaalisten palvelujen nopeaan kehitykseen ja verkottuneeseen tiedon käyttämiseen. Erilaisista lähteistä kerättävän ja tallennettavan raakadatan hyödyntäminen on vasta aluillaan, joten aihepiiristä kiinnostuneille löytyy uusia mielenkiintoisia tutkimusaiheita jatkossakin.

LÄHTEET

Avoindata.fi. 2014. Hakupäivä 4.12.2014, <https://www.avoindata.fi/data/fi/dataset>.

Creative Commons Suomi. 2014. CC-lisenssin merkitsemisohjeet. Hakupäivä 23.9.2014, <http://creativecommons.fi/lisenssit/lisenssien-kayttoohje/cc-lisenssin-merkitsemisohjeet/#more-429>.

Dietrich, D., Gray, J., McNamara, T., Poikola, A., Pollock, R., Tait, J. & Zijlstra, T. 2012. Open Data Handbook. Open Knowledge Foundation. Hakupäivä 23.1.2014, <http://opendatahandbook.org/pdf/OpenDataHandbook.pdf>.

Forum Virium Helsinki. 2014. Apps4Finland seuloo avoimen datan helmet. Hakupäivä 13.3.2014, <http://www.forumvirium.fi/node/1488>.

Helsinki Region Infoshare. 2014. HRI-palvelu. Hakupäivä 14.3.2014, <http://www.hri.fi/fi/hri-projekti/>.

Henkilötietolaki 22.4.1999/523.

JUHTA - Julkisen hallinnon tietohallinnon neuvottelukunta. 2014. JHS XXX Avoimen tieto-aineiston käyttöluupa. Versio 0.14, 28.02.2014. Hakupäivä 18.9.2014, http://www.jhs-suositukset.fi/c/document_library/get_file?uuid=59208dbd-e396-495d-a838-b7ca01c7f4af&groupId=14.

Kaario, K. & Peltola, T. 2008. Tiedonhallinta : avain tietotyön tuottavuuteen. Porvoo: WS Bookwell.

Kaleva. 2013. Oulu avasi tietoaineistoja vapaaseen käyttöön. Hakupäivä 5.12.2014, <http://www.kaleva.fi/uutiset/oulu/oulu-avasi-tietoaineistoja-vapaaseen-kayttoon/649055/>.

Laki viranomaisten toiminnan julkisuudesta 21.5.1999/621.

OKFN LOCAL: Finland. 2014. Open Knowledge Roadshow 2014. Hakupäivä 5.12.2014, <http://fi.okfn.org/roadshow2014/>.

Open Knowledge Finland. 2013. OKFFI Roadshow 2013: Oulun avoin data rajapinta auki. Video. Hakupäivä 5.12.2014, www.youtube.com/watch?v=dFbijSUPh_s.

Partanen, P & Upola T. 2013. Helsinki Region Infoshare - 2 vuotta avointa julkista dataa. Virikkeitä tietovarantojen avaajille. Helsinki: Helsinki Region Infoshare. Hakupäivä 31.1.2014, http://www.hri.fi/2vuotta/HRI-2vuotta_web.pdf.

Poikola, A., Kola, P. & Hintikka, K. A. 2010. Julkinen data - johdatus tietovarantojen avaamiseen. Helsinki: Liikenne- ja viestintäministeriö. Hakupäivä 15.1.2014, <http://www.julkinendata.fi/>.

Salokannel, M. 2012. Julkisesta datasta avoimeen dataan: julkisen datan lisensiointi. ETLA Keskusteluaiheita nro 1277. Hakupäivä 6.3.2014, <http://www.etla.fi/wp-content/uploads/2012/09/dp1277.pdf>.

Sanastokeskus TSK ry. 2013. Tietotekniikan termitalkoot. Hakupäivä 28.2.2014, <http://www.tsk.fi/tsk/termitalkoot/>.

Tietosuojavaltuutetun toimisto. 2014. Lait. Hakupäivä 8.4.2014, <http://www.tietosuoja.fi/1556.htm>.

Toikkanen, T. 2014. Tekijänoikeudet ja creative commons datan avaajalle. Hakupäivä 3.10.2014, <http://www.slideshare.net/tarmot/tekijanoikeudet-ja-creative-commons-datan-avaajalle>.

Valtion maksuperustelaki 21.2.1992/150.

Valtioneuvosto. 2011. Valtioneuvoston periaatepäätös julkisen sektorin digitaalisten tietoaaineistojen saatavuuden parantamisesta ja uudelleenkäytön edistämisestä 3.3.2011. Hakupäivä 16.3.2014, [http://www.lvm.fi/c/document_library/get_file?folderId=1551281&name=DLFE-11828.pdf&title=Ehdotus valtioneuvoston periaatepaatokseksi - Julkinen tietoaaineisto \(3.3.2011\).pdf](http://www.lvm.fi/c/document_library/get_file?folderId=1551281&name=DLFE-11828.pdf&title=Ehdotus+valtioneuvoston+periaatepaatokseksi+-+Julkinen+tietoaaineisto+(3.3.2011).pdf).

Valtiotason arkkitehtuurit -hanke. 2010. Julkisen hallinnon tietoarkkitehtuuri. Määrittely 1.00 12.11.2010. Valtiovarainministeriö. Hakupäivä 8.3.2014, http://www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/03_muut_asiakirjat/20100105Valtio/05_Julkishallinnon_tietoarkkitehtuuri.pdf.

Valtiovarainministeriö. 2013. Päätös avoimen tiedon ohjelmasta. Asettamispäätös VM041:00/2013. Hakupäivä 17.3.2014, http://www.vm.fi/vm/fi/03_tiedotteet_ja_puheet/01_tiedotteet/20130517ietova/Asettamispaeaetoes_avoin_tieto.pdf.

Valtiovarainministeriö. 2014. Avoimen tiedon ohjelma. Hakupäivä 4.12.2014, http://www.vm.fi/vm/fi/05_hankkeet/02381_avoin_tieto/index.jsp.