

Jerkko Hietanen

AVOIMEN TIEDON STRATEGIAT JA KÄYTÄNNÖT

Vertailu avoimen tiedon jakelusta kolmessa maassa

TIIVISTELMÄ

Jerkko Hietanen: Avoimen tiedon strategiat ja käytännöt – vertailu avoimen tiedon jakelusta kolmessa maassa
Kandidaatintutkielma
Tampereen yliopisto
Infomaatiotutkimus ja interaktiivinen media

Toukokuu 2020

Tutkimuksessa selvitetään julkisten tietovarantojen avaamisen periaatteita ja käytännön toimenpiteitä. Tutkimuksen kohteena ovat kolmen valtion, Suomen, Tanskan ja Viron, kansalliset strategiat ja toimenpiteet avoimen tiedon julkistamisessa. Avoimen tiedon julkistamisen periaatteita tarkastellaan sisällönanalyysin menetelmiä käyttäen. Analysoitavana ovat viralliset avoimen tiedon strategiat tai julkiset linjaukset, joita viranomaiset ovat tehneet. Sisällönanalyysissä arvioidaan varsinaisen tekstin analysoinnin rinnalla strategioiden taustalla olevia tavoitteita ja arvoja.

Tietoaineistojen julkistamisen arvioinnissa hyödynnetään kansallisten avoimen tiedon portaalien lisäksi eurooppalaista avoimen tiedon portaalia. Portaalien sisältämiä tietoja verrataan avoimen tiedon julkaisemisen kriteereihin. Avoimen tiedon piiriin on eri maissa sisällytetty hyvinkin erilaisia tietoaineistoja. Tietoaineistojen julkaisemista avoimena tietona on analysoitu tarkemmin kahdella aihealueella (liikennetieto ja ympäristötieto).

Avoimen tiedon strategioita ja käytännön jakelua koskevan tutkimuksen keskeiset tulokset koskevat ensinnäkin eroja kansallisten avoimen tiedon strategioiden välillä ja toiseksi eroja avoimen tiedon jakelussa ja etenkin yhdistetyn avoimen tiedon saatavuudessa. Avoimen tiedon strategioissa ovat selkeät erot avoimen tiedon hallinnoissa sekä avoimen tiedon standardien hyödyntämisessä. Virossa valtion rooli on keskeisempi kuin Suomessa tai Tanskassa. Kaikkiaan avoimen tiedon strategiat ovat Suomessa, Tanskassa ja Virossa sisällöltään eri tavoin painottuneita. Avoimen tiedon jakelussa erot Tanskan, Suomen ja Viron välillä ovat merkittäviä. Tanskassa ja Suomessa on käytössä kehittyneitä avoimen tiedon tiedostomuotoja, mutta Virossa avoin tieto julkaistaan pääosin pelkistettyinä verkkosivuina.

Tutkimuksessa havaittiin, että Suomessa avoimen tiedon jakelu on monipuolisempaa ja laajempaa kuin Tanskassa tai Virossa. Analysoitavina olleiden ympäristötiedon ja liikennetiedon osalta Suomessa on tarjolla merkittävästi enemmän aineistoja kuin vertailumaissa.

Avainsanat: avoin tieto, julkinen tieto, viranomaiset, kansalliset strategiat, portaalit

Sisällysluettelo

1	JOHDANTO	2
2	KESKEISISTÄ KÄSITTEISTÄ.....	4
2.1	Avoimen tiedon käsitteistä	4
2.2	Avoimen tiedon strategia	5
2.3	Yhdistetyn avoimen tiedon viiden tähden kriteerit	6
2.4	Yhdistetyn avoimen tiedon seitsemän tähden kriteerit	7
2.5	Tanskan viiden tähden perustietovarantojen kriteerit.....	7
2.6	Avoin tieto, julkisen tiedon uudelleenkäyttö ja yksityisyydensuoja	7
3	AIHEPIIRIÄ KOSKEVIA AIEMPIÄ TUTKIMUKSIA JA SELVITYKSIÄ.....	11
3.1	Huijboomin ja Van Den Broekin tutkimus avoimen tiedon strategioista	11
3.2	Kassenin tutkimukset avoimen tiedon linjauksista ja toteutuksesta.....	12
3.3	Euroopan komission tutkimus ”Open Data Maturity”.....	12
3.4	Muita aihepiiriä koskevia tutkimuksia.....	14
4	TUTKIMUSMENETELMISTÄ JA TUTKIMUSAINEISTOSTA.....	15
5	SISÄLLÖNANALYYSI – AVOIMEN TIEDON STRATEGIAT.....	17
5.1	Avoimen tiedon strategiat – sisällönanalyysi.....	17
5.1.1	Tanska.....	17
5.1.2	Viro.....	18
5.1.3	Suomi.....	20
5.1.4	Avoimen tiedon strategioiden vertailua	22
6	AVOIMEN TIEDON SAATAVUUS – ANALYYSI.....	24
6.1.1	Avoimen tiedon aineistojen analyysistä	24
6.1.2	Tanska.....	24
6.1.3	Viro.....	25
6.1.4	Suomi.....	25
6.1.5	Avoimen tiedon saatavuus - vertailua	26
7	TUTKIMUSTULOKSET JA PÄÄTELMÄT	28
	LÄHTEET	29

1 JOHDANTO

Tietoyhteiskuntaa koskevaan julkiseen keskusteluun on viimeksi kuluneiden kymmenen vuoden aikana tullut uutena teemana avoin julkinen tieto. Avoin julkinen tieto tarkoittaa käytännössä julkisen hallinnon tietovarantojen avaamista sekä kansalaisten että yritysten vapaaseen käyttöön helposti hyödynnettävässä muodossa. Suomessa merkittävä linjaus tehtiin vuonna 2011 valtioneuvoston periaatepäätöksessä, joka koski julkisen hallinnon digitaalisten tietoaineistojen saatavuutta. Periaatepäätöksen mukaan "tietoaineistojen tulee olla avoimesti saatavilla ja uudelleenkäytettävissä yhteinäisin, selkein ja kaikille tasapuolisin ehdoin, pääsääntöisesti maksutta." (VN periaatepäätös 03.03.2011). Julkisten tietovarantojen avaamista koskevaa keskustelua oli toki käyty jo ennen periaatepäätöstä (Poikola et al. 2010, 16-19).

Tämän tutkimuksen tavoitteena on tutkia avoimen tiedon linjauksia kolmessa Euroopan valtiossa sekä selvittää, miten ylätasoinen linjaukset ovat käytännössä havaittavissa avoimen julkisen tiedon jakelussa. Erityiskysymyksenä kiinnostaa avoimen tiedon julkaisemisen kriteeristö, sillä avoimen tiedon piiriin on eri maissa sisällytetty hyvinkin erilaisia tietoaineistoja. Yhdistetyn avoimen tiedon (linkitetyn avoimen datan/ linked open data) kriteerit ovat yksi tarkastelun kohde. Yhdistetyn avoimen tiedon viiden tähden laatukriteerien (Berners-Lee 2012) lisäksi käsiteltävänä on laajempi seitsemän tähden laatukriteeristö (Hyvönen et. al 2014). Tarkasteluun on Suomen lisäksi valittu Tanskan ja Viron avoimen tiedon jakelun kansalliset linjaukset ja toimintatavat. Tanskassa ja Virossa on jo pitkään korostettu digitalisaation ja avoimen tiedon merkitystä ja kumpikin maa on viime vuosina edennyt erilaisten sähköisten palvelujen käyttöönotossa. Sen vuoksi ne ovat mielenkiintoinen vertailukohta Suomelle. Euroopan unionissa avointa tietoa koskevia linjauksia on tehty avoimen tiedon direktiiveissä viimeksi vuonna 2019 ja vuoden 2020 alussa julkaistussa Euroopan datastrategiassa.

Tutkielman keskeiset kysymykset ovat:

- Millaisia kansallisen tason strategisia linjauksia (strategiat / toimintaohjelmat) on julkaistu avoimen tiedon jakelun periaatteista ja toimintatavoista sekä miltä osin kansalliset strategiat poikkeavat toisistaan tai Euroopan unionin yhteisistä linjauksista?
- Millaista avointa tietoa tutkimuksen kohteena olevissa maissa tarjotaan ympäristötiedon ja liikennetiedon aihealueilla (tarkentavana tutkimuskysymyksenä selvitetään, missä muodossa

avointa tietoa on tarjolla ja hyödynnetäänkö niissä kehittyneimpiä tiedostomuotoja, jotka mahdollistavat aineiston monipuolisen uudelleenkäytön).

Tutkielma jakautuu kuuteen lukuun, joista ensimmäinen on johdanto. Toisessa luvussa käsitellään keskeisiä käsitteitä ja kolmannessa luvussa arvioidaan aihepiiriä koskevia aiempia tutkimuksia ja selvityksiä. Neljännessä luvussa käydään läpi tutkimuksessa käytetyt analyysimenetelmät sekä tutkimusaineistot. Viidennessä luvussa käsitellään strategioiden ja käytännön tiedonjakelun vertailua kolmen maan tietoaaineistojen osalta. Viimeisessä luvussa esitellään tutkimuksen tulokset ja päätelmät.

2 KESKEISISTÄ KÄSITTEISTÄ

2.1 Avoimen tiedon käsitteistä

Avoimen tiedon julkaisemiseen liittyy vielä nykyäänkin selkiytymätön ja epätarkka käsitteiden käyttö. Suomessa on Kansalliskirjaston Finto –asiasanastoon ja ontologiapalveluun sisältyvässä YSO-ontologiassa eli yleisessä suomalaisessa ontologiassa onnistuneesti tehty ero keskeisten käsitteiden välillä. Avoin tieto on ontologian mukaan yläkäsite, jonka alakäsitteenä on yhdistetty avoin tieto. Yhdistetty avoin tieto puolestaan korvaa käytössä olleet termit avoin linkitetty tieto, avoin yhdistetty tieto, linked open data, linkitetty avoin tieto ja open linked data (<http://www.yso.fi/onto/yso/p26001>).

Avoin tieto ja yhdistetty tieto

Avoin tieto tarkoittaa pelkistetysti määriteltynä tietoa, joka on vapaasti käytettävissä ja kierrätettävissä avoimella käyttöoikeuslisenssillä. ”Avoin data on digitaalisessa muodossa olevaa informaatiota, joka on kaikkien vapaasti käytettävissä mihin tahansa käyttöön, kunhan sen alkuperäinen lähde mainitaan” (Avoindata.fi 2020). Yhdistetty tieto tai linkitetty tieto tarkoittaa puolestaan tietoelementtejä, jotka on linkitetty toisiinsa. Yhdistetty tai linkitetty tieto voidaan nähdä kehitysaskelena internetin kehityskulussa, jossa ovat yksittäisinä askeleina linkitetyt palvelimet, linkitetyt dokumentit ja linkitetyt tiedot.

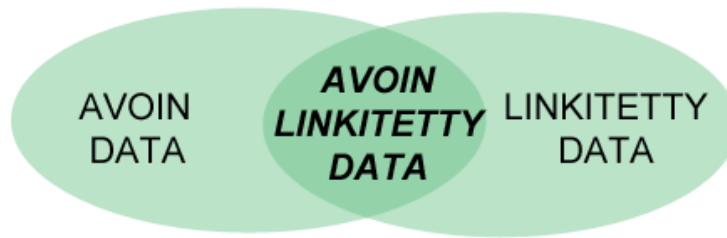
Yhdistetty avoin tieto

Yhdistetty avoin tieto tarkoittaa tietoa, joka on avattu sellaisessa muodossa, että myös muut järjestelmät pystyvät hyödyntämään sitä. Yhdistettyä avointa tietoa voidaan laajentaa, kierrättää ja yhdistellä edelleen eri sovelluksissa.

Heinemannin (2016) määritelmien mukaan

- Avoin tieto (Open Data) tarkoittaa yksinkertaisesti digitaalisessa muodossa olevaa tietoa, joka on vapaasti käytettävissä ja kierrätettävissä avoimella lisenssillä.
- Yhdistetty (linkitetty) tieto (Linked Data) puolestaan tarkoittaa tietoelementtejä, jotka on linkitetty toisiinsa. Se on seuraava askel internetin loogisessa jatkumossa linkitetyt palvelimet > linkitetyt dokumentit > linkitetyt tiedot.

- Avoin yhdistetty (linkitetty) tieto (Linked Open Data) tarkoittaa dataa, joka on avattu sellaisessa muodossa, että myös muut järjestelmät pystyvät hyödyntämään sitä. Avointa linkitettyä dataa voidaan laajentaa, kierrättää ja yhdistellä edelleen eri sovelluksissa.



Kuvio 1: Avoimen tiedon, linkitetyn tiedon ja avoimen linkitetyn tiedon välinen suhde (Heinemann 2010, 1)

Avoin yhdistetty tieto liittyy vahvasti semanttisen verkon (semantic web) standardien kehittämiseen. Semanttisen verkon tavoitteena on saada tietosisällöt julkaistua koneluettavassa muodossa eli tavalla, jota myös koneet ymmärtävät. Semanttisen verkon keskeisiä standardeja ovat W3C-organisaation (World Wide Web Consortium) julkaisemat RDF-tietomallistandardi (Resource Description Framework) ja sitä hyödyntävä SPARQL-kyselykieli (Hyvönen 2018, 24-27; Lohtander 2014, 16-19).

2.2 Avoimen tiedon strategia

Avoimen tiedon strategia tarkoittaa poliittista linjausta siitä, miten julkisia tietovarantoja avataan ja miten laajalti tietoa tarjotaan yritysten ja kansalaisten hyödynnettäväksi.

Avoimen tiedon julkaisemisessa noudatetaan useita periaatteita, jotka voidaan tiivistää seuraaviin (Open Government Data 2007):

- Avoimena tietona julkaistava tieto on julkista tietoa. Tämä käytännössä tarkoittaa sitä, että henkilötietojen suoja ja yksityisyyden suoja on taattava, kun tietoa julkistetaan.
- Tiedon on oltava koneellisesti käsiteltävässä muodossa, jotta sitä voivat käsitellä erilaiset sovellukset ja tietoa voidaan välittää erilaisten rajapintojen kautta.
- Tietoa voidaan käyttää maksutta.
- Tiedon julkaisija (tuottaja) mahdollistaa tiedon uudelleenkäytön ja jatkojalostamisen selkeillä käyttöehdoilla

- Tiedon rakenne on kuvattu selkeästi ja metatietostandardien mukaisesti.

Kansalliskirjaston muotoilemat periaatteet avoimen tiedon saatavuudesta ovat hieman laiveampia (Kansalliskirjasto 2011):

- **Laillisuus** - avoin saatavuus ottaa huomioon tekijänoikeudet, yksityisyyden suojan ja muut säännökset.
- **Tuoreus**. Avoin saatavuus on toteutettava mahdollisimman nopeasti aineiston valmistuttua.
- **Löydettävyyys**. Avoimesti saatavan aineiston tulee olla mahdollisimman helposti löydettävissä.
- **Helppokäyttöisyys**. Avoimeen saatavuuteen kuuluu aineiston hyvä ja esteetön käytettävyys.
- **Pysyvyys**. Avoin saatavuus on pysyvää; aineistojen pysyminen saatavana tulee suunnitella tarvittaessa useidenkin vuosien aikajänteellä.
- **Kuvailutiedon avoimuus**. Avoimesti saatavan aineiston kuvailutiedot ovat myös avoimesti saatavilla.

2.3 Yhdistetyn avoimen tiedon viiden tähden kriteerit

Internetin www-käyttöympäristön perustajaksi mainittu Tim Berners-Lee on kehittänyt yhdistetyn avoimen tiedon viiden tähden kriteerit (Berners-Lee 2012, 1). Berners-Leen kriteerit ovat luonteeltaan hierarkkisia eli kahden tähden taso sisältää yhden tähden tason ominaisuudet.

Berners-Leen viisiportainen luokitus sisältää seuraavat kriteerit:

* saatavilla verkossa (missä tahansa tiedostomuodossa) avoimella lisenssillä

** saatavilla koneluettavassa rakenteisessa muodossa (esimerkiksi excel-taulukkona eikä vain kuvatie-dostona)

*** kuten ** mutta järjestelmäriippumattomassa formaatissa (esim. CSVnä eikä excelinä)

**** kaikki edellä mainitut ja lisäksi: käytetty W3C:n avoimia standardeja (RDF ja SPARQL) identifioimaan tietoja, niin että julkaistuun tietoon voidaan viitata

***** kaikki edellä mainitut ja lisäksi: linkitää tietosi muiden tietoihin liittääksesi sen oikeaan kontekstiin.

2.4 Yhdistetyn avoimen tiedon seitsemän tähden kriteerit

Professori Eero Hyvönen on tutkijaryhmänsä kanssa kehittänyt Berners-Leen kriteerejä laajemman tähdistön (Hyvönen 2014). Hyvösen seitsemän tähden kriteerit perustuvat viiden ensimmäisen kriteerin osalta Berners-Leen luokitukseen. Näiden lisäksi Hyvönen on julkaissut kaksi uutta ylempää kriteeriä ja itse asiassa Hyvönen kutsuu seitsemän tähden malliaan linkitetyn datan palvelumalliksi. Tähdet ovat ikään kuin kannustetähtiä, joiden avulla halutaan kiinnittää huomiota avoimen tiedon julkaisemisen kannalta keskeisiin asioihin. Kuudes tähti kuvaa sitä, miten hyvin datan skeemat (eli se, mitä data on) on julkaistu. Seitsemäs tähti myönnetään, mikäli data täyttää myös tähtiportaikon laatuvaatimukset ja mikäli aineiston tuottaminen on riittävän hyvin dokumentoitu.

2.5 Tanskan viiden tähden perustietovarantojen kriteerit

Tanskassa on otettu käyttöön viiden tähden luokitus perustietovarantojen laatukriteeriluokitus (Røyen 2014). Viiden tähden luokitus ei ole samalla tavalla hierarkkinen kuin Berners-Leen tai Hyvösen kriteeristö, vaan se on pikemminkin periaatteita koskeva minimivaatimus avoimen tiedon julkistamiselle.

Tanskan viiden tähden luokitus koostuu seuraavista osista (Røyen 2014):

- * perustietovarantojen on oltava riittävän laadukkaita suurimpaan osaan käyttötarkoituksista.
- * perustietovarantojen oikeellisuudesta ja ajantasaisuudesta huolehtiminen on oltava selkeästi vastuutettu ja tehokkaasti hoidettu.
- * perustietovarannot ovat semanttisesti yhteensopivia ja ne on mallinnettu "Perustietojen mallinussääntöjen" mukaan
- * perustietovarannot ovat vapaasti saatavilla ja niiden uudelleenkäyttöä ei ole rajoitettu (ei koske arkaluontoista tietoa), tiedot ovat saatavissa tiedonjakelijan kautta
- * julkinen sektori ja yksityiset käyttävät perustietovarantoja, ja niiden käytöstä saadaan merkittävää hyötyä

2.6 Avoin tieto, julkisen tiedon uudelleenkäyttö ja yksityisyydensuoja

Avoimen tiedon julkaisemisen periaatteisiin on vuodesta 2003 alkaen vaikuttanut EU:n julkisen sektorin tiedon uudelleenkäyttöä koskeva PSI-direktiivi (Knuuti 2014, 52-53). Direktiivin tavoit-

teena on edistää Euroopan unionissa julkisen sektorin hallinnassa olevien tietovarantojen hyödyntämistä ja tietoaineistojen jatkojalostusta. Direktiivin taustalla olivat arviot siitä, että EU:n ulkopuoliset valtiot, etenkin Yhdysvallat ja Japani, saivat kilpailuetua siitä, että ne hyödynsivät yritystoiminnassa laajasti julkisen sektorin tuottamia tietoaineistoja. Direktiivin vaikutuksia arvioitiin vuonna 2011 (Euroopan komissio 2011,). Tuolloin kävi ilmi, että direktiivin tekstimuotoilu avoimen tiedon julkaisemisen kohdalla oli liian tulkinnanvarainen eikä direktiivi ollut merkittävästi lisännyt julkisen sektorin tiedon uudelleenkäyttöä (Attard et. al. 2016, 6).

Vuonna 2013 direktiiviä muutettiin, jotta EU:n jäsenvaltioille tulisi mahdollisimman selkeä velvollisuus sallia kaikkien tietoaineistojen uudelleenkäyttö, jollei niiden saatavuutta ole rajoitettu tai estetty kansallisissa säädöksissä (Knuuti 2015, 55). PSI-direktiivin vaikutus ei kuitenkaan ulotu asiakirjajulkaisuutta tai yksityisyyden suojaan koskevaan kansalliseen lainsäädäntöön, joten osa julkisen sektorin tietoaineistoista (henkilötietoja sisältävät aineistot ja salassa pidettävät aineistot) jää jatkossakin direktiivin vaikutusten ulkopuolelle. Direktiivillä on silti selvä vaikutus tietoaineistojen avaamiseen, sillä se on pitänyt avoimen tiedon julkistamisen näkyvästi keskusteluteemana. Euroopan komissio on seurannut direktiivin täytäntöönpanoa erillisillä verkkosivuilla (Implementation 2017, 2). Vuonna 2019 julkaistiin PSI-direktiivi ja sen muutokset uudelleenlaadittuna, *direktiivi 2019/1024 avoimesta datasta ja julkisen sektorin hallussa olevien tietojen uudelleenkäytöstä*. EU-jäsenvaltioiden on saatettava direktiivi kansallisesti voimaan 17.7.2021 mennessä.

PSI-direktiivin täytäntöönpanossa on ollut eri EU-jäsenvaltioissa erilaisia linjauksia (Implementation 2017, 2). Useat EU-jäsenvaltiot (mm. Saksa, Italia, Espanja ja Ruotsi) ovat säätäneet erikseen julkisen sektorin tiedon uudelleenkäyttöä koskevia säädöksiä. Muutamat EU-maat ovat käyttäneet yhdisteltyjä toimenpiteitä, joihin sisältyy uutta lainsäädäntöä ja muutoksia aiempaan lainsäädäntöön (mm. Itävalta, Slovenia ja Tanska). Osa EU-maista on vain mukauttanut aiempia tiedon saatavuutta koskevia säädöksiä (mm. Alankomaat, Puola, Ranska, Suomi ja Viro). Suomessa PSI-direktiivi ja sen muutokset on implementoitu valtion maksuperustelailta (150/1992), julkisuuslailla (621/1999) sekä lailla julkisen hallinnon tiedonhallinnasta (906/2019). Virossa PSI-direktiivi ja sen muutokset on vastaavasti pantu täytäntöön Viron omalla julkisuuslailla (Avaliku teabe seadus vuodelta 2003) ja sen kahdella muutoksella vuosina 2012 ja 2015 ((Implementation 2017, s. 4).

Avoimen tiedon julkistamisen taloudellisista vaikutuksista on edelleen melko niukasti tietoa (Auer, S., Bryl, V., Tramp, S. 2014; European Data Portal 2020; Wood 2011). Suomessa on julkaistu yksi laaja selvitys avoimen tiedon (massatiedon) hyödyntämisestä suomalaisyrityksissä (Koski et. al. 2017). Siinä on pystytty osoittamaan, että yritykset, jotka käyttävät hyödyksi avointa dataa ja mas-

sadataa innovaatiotoiminnassaan, tuottavat markkinoille uusia tuoteinnovaatioita suhteellisesti huomattavasti useammin kuin yritykset, jotka hyödyntävät avointa dataa vain vähän. Tutkimuksessa todettiin, että tämä havainto piti paikkansa etenkin keskisuurissa tai suurissa teknologiayrityksissä. (Koski et. al. 2017, 23).

Eri maissa avoimen tiedon strategioihin ja tiedon käytännön jakelua koskeviin linjauksiin yleensä sisältyy näkemys siitä, miten tietoa voidaan jakaa loukkaamatta yksilöiden yksityisyyden suojaaja tai tietosuojaa (henkilötietojen suojaaja) (Simperl et. al 2016, 7). Useimmiten yksityisyyden suojaaja tai henkilötietojen suojaaja koskevat linjaukset eivät kuitenkaan täsmennä sitä, mihin aineistoihin nämä linjaukset ulottuvat. Suomessa on painotettu yksityisyyden suojaaja jo julkisten tietovarantojen avaamisen periaatteita koskevista linjauksista (Valtiovarainministeriö 2013) lähtien. Linjauksessa asetettiin tavoitteeksi, että ”kaikki merkittävät julkishallinnon hallinnoimat julkiset tiedot ovat vuoteen 2020 mennessä saatavilla koneluettavassa muodossa, maksutta ja selkein käyttöehdoin ottaen huomioon tietosuojaa, yksityisyyden suojaaja ja muut lainsäädäntöön perustuvat rajoitukset.” (Valtiovarainministeriö 2013, 1). Perusteina olivat hyödyt kansantaloudelle, läpinäkyvyys sekä parantunut tehokkuus.

Tiedon anonymisointitarpeisiin on viitattu vasta tuoreessa avoimen tiedon tuottamisen edellytyksiä koskevassa selvityksessä (Koski et al., 2017, 59). Avoimen tiedon tietovarannoista on usein vaikea hahmottaa, mitä henkilötietoja ne voivat sisältää. Etenkin viranomaispäätöksiä sisältävät aineistot olisivat tietolähteiden avaamisen kannalta hyödyllisiä, mutta niissä on usein mukana yksityishenkilöitä koskevia tietoja, joita voidaan yhdistää jostain muusta aineistosta löytyviin tietoihin. Tällaisia aineistoja ovat esimerkiksi yritystiedot, verotustiedot, terveystiedot ja sosiaaliturvaa koskevat tiedot.

Isoja datamassoja eli *big dataa* voidaan pitää erityisenä haasteena yksityisyyden suojaajalle. Big datalla yleensä tarkoitetaan valtavia, järjestelemättömiä ja hajautuneita datamassoja, joiden käsittelyyn kehitetään kaiken aikaa uusia tekniikoita. Big datan aiheuttamat ongelmat liittyvät siihen, että kenelläkään ei ole tarkkaa tietoa big datan sisältöjen yksityiskohdista eikä etenkään siitä, onko datassa mukana henkilötietoja (Hokkanen 2015, 12).

Yhtenä käytännön ratkaisuna avoimen tiedon jakelun laajentamiselle on pidetty tiedon anonymisointia tai pseudonymisointia ennen tiedon avointa jakelua. Anonymisoinnissa tai pseudonymisoinnissa on kuitenkin omat ongelmansa, sillä automaattinen anonymisointi edellyttää kehittyneiden tekstisuodattimien käyttöä. Toisaalta anonymisointi voi johtaa tietoaineiston merkityksen kutistumiseen, jos aineistosta poistetaan paljon muutakin kuin henkilöihin liittyviä tunnisteita. Li-

säksi anonymisoitua tai pseudonymisoitua dataa voidaan palauttaa eri teknologioilla palauttaa takaisin tunnistettavaan muotoon (Lubarsky 2017, 1)

3 AIHEPIIRIÄ KOSKEVIA AIEMPIÄ TUTKIMUKSIA JA SELVITYKSIÄ

Avoimen tiedon strategioita ja käytännön jakelua koskevia tutkimuksia on tehty viime vuosina muutamia. Tässä esitellään lyhyesti Huijboomin ja Van den Broekin sekä Kassenin vertailevia tutkimuksia avoimen tiedon strategioista, ja Euroopan komission teettämä tutkimus avoimen tiedon kypsyydestä (Open Data Maturity).

3.1 Huijboomin ja Van Den Broekin tutkimus avoimen tiedon strategioista

Noor Huijboom ja Tijs Van den Broek ovat tutkimuksessaan ”Open data: an international comparison of strategies” analysoineet viiden valtion avoimen tiedon strategioita (Huijboom & Van den Broek 2011). Kohteina olivat Yhdysvaltain, Ison-Britannian, Australian, Tanskan ja Espanjan hallitusten strategiat, jotka koskivat julkisen sektorin tietojen julkaisemista avoimena tietona osana avoimen hallinnon toimintaa. Huijboom ja Van den Broek arvioivat kansallisia strategioita neljästä eri näkökulmasta: ensinnäkin koulutuksen ja neuvonnan osalta, toiseksi vapaaehtoisten toimenpiteiden osalta, kolmanneksi rahoituksen ja kehitystyön osalta sekä neljänneksi lainsäädäntötoimien ja valvonnan näkökulmasta. Tutkijat päättelivät, että tutkimuskohteina olleissa viidessä maassa painotukset vaihtelivat kohtalaisen vähän. Joitakin erityispiirteitä oli jokaisessa maassa. Esimerkiksi Tanskassa korostettiin koulutusta ja neuvontaa (mm. erilaiset avoimen tiedon kokeiluportaalit, työpajat ja koulutustilaisuudet) sekä vapaaehtoistoimia (avointa tietoa tukevan ilmapiirin luominen strategioilla ja linjauksilla). Kaikissa viidessä maassa avoimen tiedon saatavuutta ja avoimeen tietoon perustuvien uusien sovellusten luomista edistettiin mm. sovelluskilpailuilla ja avoimen tiedon portaalien julkisella rahoituksella. Samalla tavoin kaikissa viidessä maassa käytettiin avoimen tiedon saatavuuden edistämiseksi lainsäädäntöä apuna, EU-maissa saatiin hyötyä julkisen sektorin tiedon PSI-direktiivin täytäntöönpanosta.

Tutkimuksessa käsiteltiin myös keskeisiä tekijöitä, jotka helpottavat tai vaikeuttavat avoimen tiedon saatavuuden edistämistä. Helpottavista tekijöistä pidettiin Yhdysvalloissa, Australiassa ja Isossa-Britanniassa tärkeinä muiden maiden linjauksia ja käytännön toteutuksia. Tanskassa pidettiin tärkeimpinä myötävaikuttajina avointa tietoa koskevia kansalaisaloitteita ja EU-lainsäädäntöä (PSI-direktiivi) (Huijboom & Van den Broek 2011). Keskeisinä esteinä pidettiin kaikissa viidessä maassa sulkeutunutta hallintokulttuuria, yksityisyyssääntöä sekä tarjolla olevan tiedon heikkoa laatua.

Merkittävänä esteenä pidettiin myös epäselvyyttä siitä, millaisia taloudellisia vaikutuksia tiedon avaamisella käytännössä on.

3.2 Kassenin tutkimukset avoimen tiedon linjauksista ja toteutuksesta

Maxat Kassen on viime vuosina tutkinut avoimen tiedon julkaisemista koskevia strategioita ja toimintatapoja useissa eri maissa, mm. Virossa, Suomessa ja kotimaassaan Kazakstanissa (Kassen 2018a ja 2018b, Kassen 2019). Kassen on tarkastellut erityisesti julkisen sektorin roolia avoimen tiedon julkaisijana sekä yksityisen sektorin ja kansalaisjärjestöjen roolia julkisen tietovarantojen avaamisen lobbaajina ja vauhdittajana. Kassen ei tutkimuksissaan ole käsitellyt yksittäisten tietovarantojen avaamista, vaan hän on keskittynyt avoimen tiedon julkaisemiseen liittyvään vuoropuheluun julkisen sektorin ja muiden toimijoiden välillä. Viron osalta Kassen pitää tutkimuksessaan merkittävänä etuna tietovarantojen avaamisessa sitä, että julkinen sektori on ottanut huomioon yksityisen sektorin ja kansalaisjärjestöjen tarpeet avointa tietoa koskevien linjausten valmistelussa. Kassen kuitenkin huomauttaa, että Virossa avoimen tiedon aineistojen uudelleenkäytöstä ei ole juuri keskusteltu. Suomen osalta Kassen arvioi, että monet eri toimijat – julkisen sektorin lisäksi - ovat edistäneet avoimen tiedon hyödyntämistarvetta koskevaa tietoisuutta. Keskeisiä toimijoita ovat Kassenin mukana olleet mm. Open Knowledge Finland, Avoin ministeriö-hanke ja Open Ecosystem Network (Kassen 2018a, 527-528). Kassen katsoo, että tiedon tehokas uudelleenkäyttö edellyttää yksityisten kehittäjäyhteisöjen aktiivisempaa panosta, jota julkisen sektorin tulisi selkeämmin tukea (Kassen 2018a, 532).

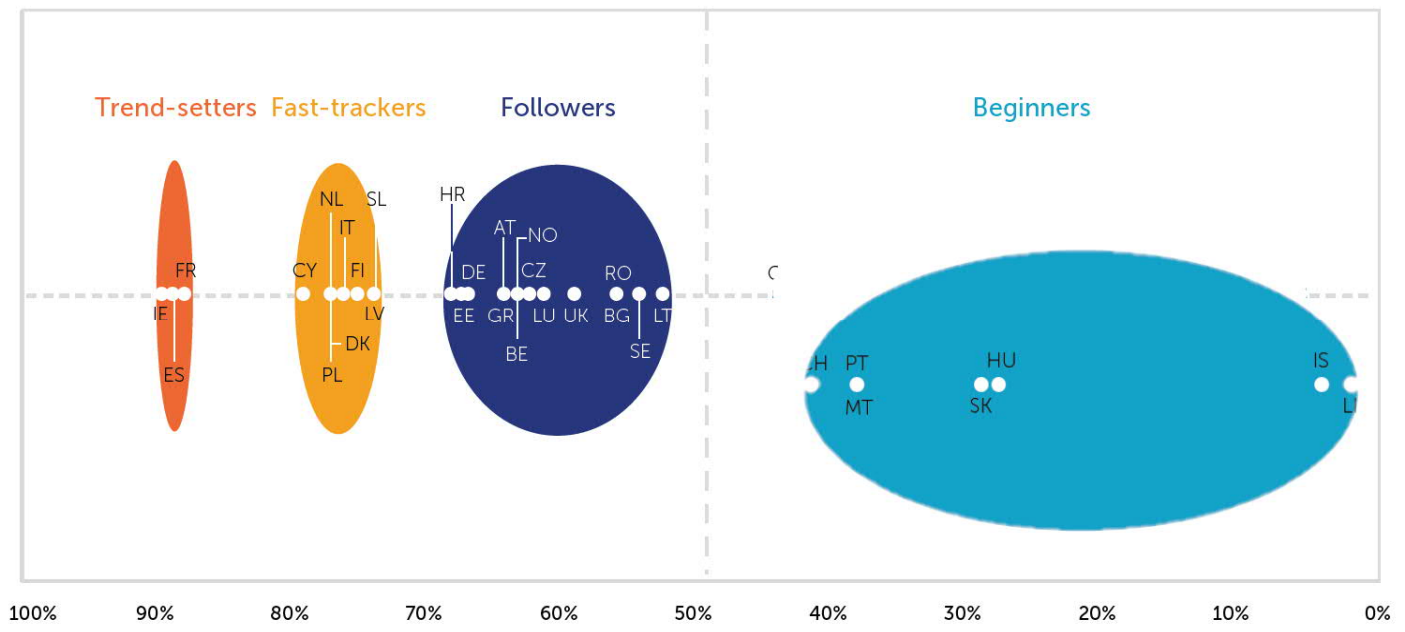
3.3 Euroopan komission tutkimus ”Open Data Maturity”

Euroopan komissio on viime vuosina teettänyt koko Euroopan unionia koskevan vuotuisen selvityksen avoimen tiedon jakelun kypsyystasosta ”Open Data Maturity”. Selvityksessä tarkastellaan avoimen tiedon saatavuutta kaikissa EU:n jäsenvaltioissa sekä Sveitsissä, Norjassa ja Lichtensteiniin. Tarkastelussa arvioidaan kahta osa-aluetta, avoimen tiedon valmiuksia (Open Data Readiness) ja avoimen tiedon portaalien kypsyyttä (Data Portal Maturity). Avoimen tiedon valmiuksien arvioinnissa on käytössä viisi kriteeriä. Ensinnäkin avointa tietoa koskevan strategian tai toimintaohjelman (policy) olemassaolo. Toisena kriteerinä ovat käyttöoikeuslisenssien linjaukset ja kolmantena kansallisen koordinaation laajuus. Neljäs kriteeri koskee tiedon käytön laajuutta ja viides koskee avoimen tiedon vaikutuksia. (Open Data Maturity 2019, s.6.)

Kypsyystasoselvityksen mukaan avoimen tiedon jakelua tukevat linjaukset ja avoimen tiedon käyttö kasvoivat merkittävästi vuodesta 2016 vuoteen 2019, samoin avoimen tiedon portaalien kypsyys (joka viittaa portaalien toimintojen monipuolistumiseen) kasvoi vuodesta 2016 noin 27%.

Open Data Maturity – selvityksessä on EU-maat jaettu neljään ryhmään. Kärkisijoilla ovat trendinmäärittäjät, joita seuraavat nopeasti etenevät maat. Niiden perässä tulevat seuraajamaat ja aloittelijat seuraavat niitä.

t



Kuvio 2: Avoimen datan kypsyys EU-maissa 2019 (Open Data Maturity 2019, 72).

Vuosien 2016 – 2019 välillä Open Data Maturity – selvityksessä sijoitukset eivät ole merkittävästi muuttuneet. Trendinmäärittäjissä on tosin enää kolme valtiota, Espanja, Irlanti ja Ranska. Suomi on pudonnut trendinmäärittäjistä nopeasti eteneviin maihin. Itävalta on pudonnut vielä selvemmin, trendinmäärittäjistä seuraajamaihin. Nopeasti eteneviin maihin kuuluvat myös mm. Italia, Latvia ja Tanska. Ruotsi ja Viro ovat pysyneet seuraajaroolissa.

Open Data Maturity –selvityksestä on syytä todeta, että arviot ovat kansallisen yhdyshenkilöverkoston tekemiä eivätkä sinänsä yhteismitallisia. Selvitys ei myöskään käsittele avoimen tiedon jakelun kansallisia linjauksia, avoimen tiedon standardien hyödyntämistä eikä esimerkiksi yhdistetyn avoimen tiedon tarjontaa kohdemaissa.

3.4 Muita aihepiiriä koskevia tutkimuksia

Taloudellisen yhteistyön ja kehityksen järjestö OECD on vuonna 2015 julkaissut raportin, jossa käsitellään Suomen ja Viron sähköisten palvelujen kehitystä ”*Estonia and Finland - Fostering Strategic Capacity across Governments and Digital Services across Borders*”. Raportissa on käsitelty sähköisen hallinnon yleisten strategioiden lisäksi myös sitä, miten julkishallinnossa hyödynnetään avointa tietoa hallinnon palvelujen kehittämiseen tai hallinnon asiakkaita koskevan tiedon keräämiseen. Raportin mukaan vain alle Virossa vain noin 25% viranomaisista ja Suomessa noin 40% viranomaisista käyttää hyödyksi avointa tietoa (OECD 2015, 139). Raportissa kuitenkin todetaan, että avoimen tiedon julkistaminen ja hyödyntäminen on etenemässä sekä Virossa että Suomessa. OECD-raportin mukaan yhtenä esteenä edelleen on, että osa viranomaisista on riippuvaisia maksullisten sähköisten palvelujen tuotoista. Toisena esteenä nähdään se, että yksityisyyden suoja aiheuttaa rajoituksia avoimen tiedon tarjonnalle.

Raportissa on keskitytty hallinnon kehitykseen ja hallinnon sähköistämiseen Virossa ja Suomessa. Raportissa ei ole mukana avoimen tiedon julkaisemista koskevien linjausten analyysia eikä raportissa arvioida lainkaan sitä, missä laajuudessa yritykset hyödyntävät avointa tietoa omassa tutkimus- ja kehitystoiminnassaan.

Euroopan ympäristövirasto EEA on teettänyt vuonna 2019 PriceWaterhouseCoopers-konsulttiyrityksellä tutkimuksen ympäristöalan avoimen tiedon saatavuudesta Euroopassa, *Open data and e-Government - good practices for fostering environmental information sharing and dissemination* (PriceWaterhouseCoopers 2019). Tutkimuksessa on käsitelty avoimen tiedon saatavuutta useassa EU-jäsenvaltiossa sekä koottu hyviä käytäntöjä ympäristötiedon avoimen jakeluun.

Suomessa on liikenne- ja viestintäministeriön työryhmä julkaissut vuonna 2013 raportin *Liikenteen ja viestinnän avoin tieto* jossa on käsitelty liikenne ja viestintäalan tietovarantoja, jotka olisi tarpeen julkistaa avoimena tietona (Liikenne- ja viestintäministeriö 2013). Raporttiin sisältyy myös tietovarantojen avaamisen tiekartta, joka monelta osin on jo toteutunut.

4 TUTKIMUSMENETELMISTÄ JA TUTKIMUSAINEISTOSTA

Avoimen tiedon strategioita ja käytännön jakelua koskevassa tutkimuksessa keskeinen tutkimusmenetelmä on sisällönanalyysi. Sisällönanalyysissä aineistoa tarkastellaan tekstisisältöä eritellen sekä etsien mahdollisia yhtäläisyyksiä ja eroja vertailun kohteina olevien tekstiaineistojen välillä. Sisällönanalyysi on tekstianalyysia, jossa tarkastellaan jo valmiiksi tekstimuotoisia tai sellaiseksi muutettuja aineistoja. Sisällönanalyysi on valittu tutkimusmenetelmäksi sen vuoksi, että sitä käyttämällä on mahdollista arvioida avointa tietoa koskevia asiakirjoja usealla eri tasolla ja useasta eri näkökulmasta.

Tutkielmassa analysoidut tekstit ovat valmiiksi tekstimuotoisia viranomaisten julkaisemia strategia-asiakirjoja sekä avoimen tiedon saatavuutta koskevia asiakirjoja ja selosteita. Sisällönanalyysin avulla pyritään muodostamaan tutkittavasta ilmiöstä tiivistetty kuvaus, joka kytkee tulokset ilmiön laajempaan kontekstiin ja aihetta koskeviin muihin tutkimustuloksiin. (Tuomi J., Sarajärvi 2002, 105). Krippendorff korostaa, että sisällönanalyysi ei ole pelkkää tekstin välittämän viestin analyysiä (kuka viestii mitä kenelle, millä välineellä, millä tavoitteilla), vaan samalla täytyy arvioida myös psykologisia näkökohtia (miksi viestitään tietyllä tavalla), institutionaalisia näkökohtia (mitkä arvot ovat viestin takana) sekä kulttuurisia näkökohtia (mitkä rituaalit tai perinteet ovat viestissä mukana) (Krippendorff 2012). Etenkin avoimen tiedon strategioiden (jotka ovat poliittisen päätöksenteon tuotoksia) sisällönanalyysissä on tärkeä ottaa huomioon institutionaaliset ja kulttuuriset näkökohdat.

Bengtsson (Bengtsson 2016, 9) erottaa sisällönanalyysin kvalitatiivisissa menetelmissä lausuma- tai manifestitason analyysin (manifest analysis) ja latenttialyysin (latent analysis). Ensimmäisessä tutkitaan pintarakennetta eli mitä tekstissä on kerrottu ja toisessa vastaavasti syvärakennetta eli mitä on ollut tarkoitus kertoa. Pintarakennetta ja analysoidavaa tekstiä pitäisi pystyä tarkastelemaan myös pintaa syvemältä ja arvioida, mitkä ovat olleet tekstin tavoitteet. Samaan seikkaan viittaa myös Mayring (2014, 39), jonka mukaan sisällönanalyysissä on korostettava analysoidavan aineiston viestinnällistä kontekstia eli mihin asiayhteyteen kyseinen teksti liittyy. Sisältöä analysoitaessa pitää ottaa huomioon tekstin alkuperä ja tekstillä tavoitellut vaikutukset. Sisällönanalyysiä menetelmänä on myös kritisoitu tekstien tulkinnan uskottavuuden ja analyysiprosessin osalta (Elo et al. 2014, 2).

Avoimen tiedon jakelua ja saatavuutta koskevassa selvityksessä sisällönanalyysiä käytetään yhdistettynä tilastollisiin menetelmiin. Sisällönanalyysissä arvioin yhtäältä strategisten linjausten lausu-

matasoa (manifestitasoa) ja toisaalta latenttia tasoa. Sisällönanalyysin kohteina käytän olemassa olevia avoimen tiedon portaaleja, suomalaista avoindata.fi -portaalia, tanskalaista opendata.dk -portaalia, virolaista opendata.riik.ee -portaalia sekä eurooppalaista dataportaalia (<https://europeandataportal.eu/>). Kokoan niistä tiedot, jotka koskevat käsittelemiäni aihealueiden tietoaaineistoja. Menetelmänä on tekstianalyttinen ja osittain tilastollinen vertailu siitä, mitä aihealueiden tietoaaineistoja olemassa, millaisessa muodossa tarjolla, mitkä ovat käyttöoikeudet sekä analyttinen laadullinen arviointi siitä, miten tietoja voidaan yhdistetyn avoimen tiedon laatukriteeristöissä.

5 SISÄLLÖNANALYYSI – AVOIMEN TIEDON STRATEGIAT

Avoimen tiedon strategioita ja käytännön jakelua koskevassa sisällönanalyysissä käsitellään aluksi avoimen tiedon strategioita ja toimintaohjelmia. Niiden jälkeen siirrytään avoimen tiedon jakeluun ja saatavuuteen tutkimuksen kohteina olevissa maissa eli Tanskassa, Virossa ja Suomessa.

5.1 Avoimen tiedon strategiat – sisällönanalyysi

5.1.1 Tanska

Tanskassa on julkaistu sähköisen hallinnon strategioita (e-Government strategy) vuodesta 2001 lähtien. Strategioissa painotukset ovat muuttuneet vuosien varrella. Ensimmäisissä strategioissa korostettiin sähköistä asiointia ja sähköisten maksutapojen käyttöönottoa. Vasta vuonna 2011 strategiaan sisällytettiin tiedon uudelleenkäyttöä koskevia linjauksia. Vuodesta 2012 lähtien Tanskassa on toteutettu perustietovarantojen avaamiseen tähtäävää toimintaohjelmaa (Grunddataprogrammet) (Horst 2014, s. 1). Tanskan perustietovarantojen avaamista on perusteltu erityisesti talouskasvun vauhdittamisella, innovaatioiden edistämällä ja työllisyyden lisäämisellä. Perustietovarannoista on avattu vuodesta 2012 lähtien mm. kiinteistötiedot, osoitetiedot, yritysrekisteritiedot. Käynnissä on lisäksi mm. geotiedon (kartta- ja maaperätiedot), vesiresursseja koskevan datan sekä maatalouden tietovarantojen avaaminen.

Perustietovarantojen saatavuus on Tanskassa varmistettu keskitetyn jakelumekanismien eli Datafordeler –palvelun www.datafordeler.dk avulla. Tanskassa lähivuosien keskeisenä kehittämistavoitteena on perustietovarantojen laadun kehittäminen, jotta tietoja voidaan entistä paremmin hyödyntää (Digitaliseringsstyrelsen 2016, 27).

Tanskan avoimen tiedon strategiset linjaukset sisältyvät vuosien 2016-2020 digitalisaatiostrategiaan. Strategiassa korostetaan kolmea päätavoitetta. Ensimmäinen digitaalisen tiedon ja palvelujen tulee olla helppokäyttöisiä, helposti tarjolla ja korkealaatuisia. Toiseksi julkishallinnon digitaalisten aineistojen tulee parantaa talouskasvun edellytyksiä. Kolmanneksi tietoa-aineistojen digitalisaatioissa tulee ottaa huomioon tietoturvallisuus ja tiedon luotettavuus (Digitaliseringsstyrelsen 2016, 5). Digitalisaatiostrategiaan sisältyvässä perustietovarantojen avaamisessa korostetaan viiden tähden periaatteita, joita on edellä käsitelty luvussa 2.5. Julkisen sektorin tietojen avoimesta käytöstä strategiassa todetaan :

”Julkisen sektorin tiedon edistämiseksi, mukaan lukien tiedon kaupallinen käyttö, perustetaan yhteinen julkisen sektorin kumppanuus, joka koostuu Open Data DK:sta, Tanskan kunnista ja Tanskan yritysviranomaisesta (Erhvervstyrelsen). Kumppanuustaho neuvottelee yritysten ja asiantuntijoiden kanssa paremman kokonaiskuvan saamiseksi olemassa olevasta avoimesta tiedosta ja edistää pyrkimyksiä varmistaa avoimemman tiedon saatavuus.” (Digitaliseringsstyrelsen 2016, 40)

Tanskassa avointa tietoa koskevia linjauksia ei ole sisällytetty pelkästään digitalisointistrategiaan ja perustietovaranto-ohjelmaan, vaan myös muihin strategioihin, muun muassa Digitaalisen kasvun strategiaan (2018) ja kansalliseen tekoälystrategiaan (2019). Digitaalisen kasvun strategiassa ei käsitellä avoimen tiedon kysymyksiä erikseen, vaan siinä pelkästään todetaan avoimen tiedon kumppanuuden toteutus. Tekoälystrategiassa on tavoitteena julkisen sektorin tiedon saaminen tekoälysovellusten hyödynnettäväksi ja vuosien 2020-2021 aikana on tarkoitus valita viisi tietoaainetta tekoälysovelluksiin (Danish Ministry of Finance 2019, 3).

Tanskan avoimen tiedon strategian linjauksia voidaan sisällönanalyysissä tarkastella lausumatasolla ja latenssitasolla ((Digitaliseringsstyrelsen 2016, 40)

Lausumataso	Latentti taso
”Kumppanuus neuvottelee yritysten ja asiantuntijoiden kanssa paremman kokonaiskuvan saamiseksi olemassa olevasta avoimesta tiedosta ja edistää pyrkimyksiä varmistaa avoimemman tiedon saatavuus”	Tietojen julkaisemiseen avoimena tietona ei aseteta ennakolta erityisiä vaatimuksia. Uusia avoimen tiedon julkaisemisen kohdealueita ei ole tarpeen määritellä.

5.1.2 Viro

Virossa on yleisiä digitalisaation tavoitteita linjattu Digitaalinen Agenda 2020 –asiakirjassa ja e-Estonia-hankkeessa (e-Estonia.com). Virossa on kehitetty lukuisia viranomaisasioinnin palveluja (mm. sähköinen kaupparekisteri, väestörekisteri, kiinteistörekisteri ja potilasrekisteri), mutta avoimen tiedon julkaisemisessa Viro ei ole edennyt samaan tahtiin. Tähän paradoksiin ovat kiinnittäneet huomiota tutkimuksissaan Kassen (2019) sekä Margetts ja Naumann (2017). Kassenin mukaan Virossa ei ole nähty tarvetta erilliselle avoimen tiedon strategialle, sillä julkisen sektorin tieto on ollut laajalti vapaasti käytettävissä jo ennen avointa tietoa koskevan keskustelun käynnistymistä Virossa (Kassen 2019, 572). Margetts ja Naumann arvioivat, että Viron oletusarvoinen digitaalisuus (digital by default) ja hallituksen vahva rooli digitalisaatiossa ovat jättäneet varjoonsa avoimen tiedon kokeilukulttuuriin ja mahdollisuudet omaehtoiseen tiedonlouhintaan avoimen tiedon tietoaainetoista (Margetts & Naumann 2017, 29)

Avoimen tiedon sivurooli näkyy myös avoimena tietona julkaistujen aineistojen määrissä. Esimerkiksi Euroopan dataportaalissa on Virossa saatuja aineistoja mukana 433 tietoaaineistoa, kun Tanskasta on 1871 ja Suomesta 5050 tietoaaineistoa (European Data Portal 2020). Virossa avoimen tiedon portaali opendata.riik.ee avattiin kesällä 2018 ja se sisältää 570 tietoaaineistoa. Aineisto on siis laajempi kuin eurooppalaisen dataportaalin kautta saatava kokonaisuus.

Avoimen tiedon julkaisemisesta ja hyödyntämisestä julkaistiin Virossa ns. Vihreä kirja vuonna 2014 ja Digitaalinen agenda 2020-linjaus vuonna 2014 (Estonian Digital Agenda 2020). Vihreässä kirjassa on useita toimenpide-ehdotuksia avoimen tiedon saatavuuden parantamiseksi. Näistä suurin osa on toteutettu tai käynnistetty vuosina 2015-2020 osana e-Estonia-kehitystyötä. Avoimen tiedon palveluina on toteutettu mm. Meieraha (valtion budjetin seurantaväline) ja Valitsemine Valvurid (kansanedustajien antamien lupauksen seurantapalvelu). Virossa avoimen tiedon julkistamista koskevat tavoitteet liittyvät erityisesti avoimuuden ja läpinäkyvyyden lisäämiseen, tiedon saatavuuden parantamiseen, tietopyyntöjen määrän vähentämiseen hallinnossa, sekä uusien teknologioiden kuten yhdistetyn tiedon, Big Datan ja esineiden internetin (IoT) hyödyntämiseen (Open Data Portal of Estonia 2020, 1). Digitaalinen agenda 2020 –linjauksissa on lyhyesti mainittu avoin tieto ja yhdistetty avoin tieto – niiden yhteydessä on todettu henkilötietojen suojan olevan suuri haaste (Estonian Digital Agenda 2020, 13).

Avoimen tiedon julkaisemiseen on Virossa linjattu avoimen tiedon Vihreässä kirjassa seuraavat periaatteet (Eesti Avaandmete 2016, 11):

- a) tiedon tasapuolinen saatavuus (useita formaatteja, ilman alueellisia tms. rajoituksia)
- b) tiedon kattavuus (laatukriteeri)
- c) tiedon rakenteen eheys
- d) tiedon käyttöehtojen määrittely
- e) tiedon julkaiseminen tuoreeltaan ajantasaisena
- f) tiedon anonymisointi (yksityisyyden suoja)
- g) salassapitosäädösten ottaminen huomioon (ei julkaista salassa pidettäviä aineistoja)

Avoimen tiedon Vihreään kirjaan on lisätty muutamia lisäkriteerejä (Eesti Avaandmete 2016, 12)

”Helpon pääsyn ja mahdollisen lisäarvon saamiseksi käyttäjä on voitava

- selata ja etsiä kiinnostavien käytettävissä olevien julkisten tietojen kokoelmaa;

- ladata kokonaan tai osittain löydetty tietoaineisto hakujärjestelmän kautta välittömästi ilman, että tarvitaan sopimuksia tai salasanoja, jotka estävät välittömän lataamisen;
- käyttää tietokantaa kokonaisuudessaan vapaasti ja ladata tietokanta kokonaisuutena tietokoneelle sekä käyttää sitä omissa sovelluksissaan ilman erillistä lupaa tai (lisä) maksua.

Lisäksi Vihreän kirjan mukaan tietoa julkaisevalla julkisen sektorin laitoksella ei ole velvollisuutta tarjota tiedon käyttäjille lisäpalveluja, kuten muuntaa heille sopivaan muotoon, luoda erityisiä verkkopalveluita, kääntää aineistoja jne.

Vihreän kirjan linjauksia voidaan sisällönanalyysissä tarkastella lausumatasolla ja latenssitasolla (Eesti Avaandmete 2016, 12)

Lausumataso	Latentti taso
<p>”Tietoja julkistettaessa on etsittävä kompromissia kolmen tavoitteen välillä:</p> <ul style="list-style-type: none"> • tietojen helppo käytettävyys ja ymmärrettävyys tiedonhakijalle ja tiedon lataajalle; • tietojen julkistamisen yksinkertaisuus ja julkaisijan työvoimakustannusten minimointi; • niiden henkilöiden yksityisyyden suoja, joiden tiedot sisältyvät tietoihin.” 	<p>Tietojen julkaiseminen avoimena tietona on viranomaisen harkinnassa ja sitä voi vaikeuttaa julkaisuprosessin monimutkaisuus ja julkaisukustannukset</p>
<p>”Tietoja julkaisevalla julkisen sektorin laitoksella ei ole velvollisuutta tarjota tiedon käyttäjille lisäpalveluja, kuten muuntaa heille sopivaan muotoon”</p>	<p>Tietoja ei ole tarpeen tarjota useissa eri tiedostomuodoissa</p>

5.1.3 Suomi

Suomessa avoimen tiedon julkistamisen periaatteita ja toteutusta on linjattu *Avoimen tiedon ohjelmasssa* vuosina 2013-2015. Ohjelman käynnistymisvaiheessa valtiovarainministeriö kokosi avoimen tiedon julkistamista koskevat periaatteet (Valtiovarainministeriö 2013). Periaatemuistion mukaan tietovarantojen avaamisessa tavoitellaan hyötyjä kolmelta suunnalta. Ensinnäkin kansantaloudelliset hyödyt, jotka konkretisoituvat innovaatioiden, uudenlaisen liiketoiminnan ja tuotekehityksen kautta. Toisena näkökohtana oli avoimuuden ja läpinäkyvyyden lisääminen yhteiskunnassa, mikä puolestaan lisää mahdollisuuksia suoraan kansalaisvaikuttamiseen ja helpottaa kansalaisten ja yritysten tiedonsaantia. Kolmantena taustatekijänä nähtiin hallinnon sisäisen tuottavuuden ja tehokkuuden lisääminen, kun erilaiset tietovarannot tulevat paremmin esiin ja niiden hyödyntäminen helpottuu.

Avoimen tiedon ohjelmassa vuosina 2013-2015 koordinoitiin ja toteutettiin viranomaisten tuottaman tiedon avaamistoimia. Keskeinen osa avoimen tiedon ohjelmaa on ollut kansallisen dataportaalien käyttöönotto ja sisällön kehittäminen. Avoimen tiedon dataportaalissa www.avoindata.fi on mukana noin 1700 tietoaineistoa. Tietovarantojen avaamisen periaatteina avoimen tiedon ohjelmassa korostettiin sitä, että jokainen henkilö omistaa omat tietonsa ja että yksityisyyden suoja tulee ottaa huomioon valmisteltaessa tiedon jakelua (Jaatinen, 2014, 18).

Avoimen tiedon ohjelmaa koottiin lisäksi tietovarantojen avaamiselle yhtenäisiä käytäntöjä, joilla voidaan varmistaa tietoaineistojen ja tietojärjestelmien yhteentoimivuus. Ohjelmassa kartoitettiin kansalliset tietovarannot osana tietoarkkitehtuurityötä. Avoimen tiedon ohjelmassa määriteltiin lisäksi avoimen tiedon rajapintoja ja käyttöehtoja, näkyvimpänä ohjeistuksena julkisen hallinnon suositus Avoimen tietoaineiston käyttö lupa JHS 189 (VM Avoimen tiedon loppuraportti 2015).

Avoimen tiedon ohjelma käsitteli myös tiedon hinnoitteluperiaatteita. Lähtökohtana on, että viranomaistieto on mahdollisimman laajasti maksutta saatavilla ja tieto luovutetaan maksutta tai tiedon irrotuskustannuksilla. Kyseessä ei siis ole ehdoton maksuttomuuden vaatimus tietoluovutuksissa.

Avoimen tiedon ohjelman linjauksia voidaan sisällönanalyysissä tarkastella lausumatasolla ja latenssitasolla (Valtiovarainministeriö 2013, 2)

Lausumataso	Latentti taso
<p>Taloudellisten vaikutusten arvioinnissa kiinnitetään huomiota mm. vaikutuksiin viranomaisten maksutulojen vähenemiseen, investointitarpeisiin ja maksumuutoksella saavutettaviin yhteiskunnallisiin ja taloudellisiin hyötyihin.</p>	<p>Tietojen julkaiseminen avoimena tietona on viranomaisen harkinnassa ja sitä voi hidastaa maksutulojen väheneminen ja maksumuutoksen vaikeasti arvioitavat hyödyt.</p>
<p>Tietovarantoja avataan vaiheittain mahdollisimman laajasti ja monipuolisesti eri hallinnonaloilla sitä mukaa, kun virastojen suunnitelmat ja valmiudet ovat kunnossa. Priorisoinnissa voidaan kiinnittää huomiota arvioituihin taloudellisiin, yhteiskunnallisiin tai hallinnon tehokkuushyötyihin tai tiedon käytön tai käyttäjäkunnan laajuuteen. Yksittäisten tietosäiltöjen kysyntä ei kuitenkaan ole ennakoitavissa, ja suurimmat taloudelliset ja yhteiskunnalliset hyödyt ja vaikutukset saadaan eri tietovarantojen yhdistämisestä.</p>	<p>Tietovarantojen avaamisen laajuutta ja aikataulua ei ole tarpeen täsmentää pikaisesti. Tietovarantojen priorisointia on vaikea tehdä, kun tiedon kysyntä ei ole tarkkaan tiedossa.</p>

5.1.4 Avoimen tiedon strategioiden vertailua

Tässä alaluvussa vertaillaan Tanskan, Viron ja Suomen avoimen tiedon strategioiden sisältämiä periaatteita, vertaamalla niitä avoimen tiedon julkaisemisen yleisiin periaatteisiin ja mm. Kansalliskirjaston koostamiin periaatteisiin (luvussa 2.4). Vertailussa on otettu huomioon, että tutkittavien kolmen valtion strategiset linjaukset on kirjoitettu eri tavoin eikä niistä ole helposti löydettävissä täsmälleen samoja muotoiluja (ks. myös Kassen 2018a ja Kassen 2018b).

Kaikissa kolmessa maassa on linjattu henkilötietojen suoja ja salassapitoa koskevat periaatteet samansuuntaisesti. Samoin tiedon jakelua ja rajapintoja koskevat periaatteet ovat samankaltaisia. Eroja löytyy tiedon maksuttomuutta koskevissa linjauksissa. Virossa tiedon maksuttomuutta (tai tiedon irrottamiskustannuksia) koskevat linjaukset ovat vielä tekeillä mm. avoimen paikkatiedon tarjonnan, kaupparekisterin ja kiinteistörekisterin osalta. Tanskassa ja Suomessa paikkatietoaineistoja on maksutta saatavilla, samoin kaupparekisteritiedot. Tiedon rakennemäärittelyä koskevat linjaukset poikkeavat myös toisistaan. Tanskassa ja Suomessa korostetaan standardien käyttöä, mutta Virossa yksityiskohtaiset linjaukset ovat vielä tekemättä.

Kriteeri	Tanska	Viro	Suomi
Avoimena tietona julkaistava tieto on julkista tietoa (henkilötietojen suoja ja yksityisyyden suoja taataan, kun tieto julkistetaan).	*	*	*
Tiedon on oltava koneellisesti käsiteltävissä muodossa, jotta sitä voivat käsitellä erilaiset sovellukset ja tietoa voidaan välittää erilaisten rajapintojen kautta.	*	*	*
Tietoa voidaan käyttää maksutta tai korkeintaan tiedon irrottamiskustannuksista voidaan veloittaa	*	- (linjaukset täsmentämättä)	*
Tiedon julkaisija (tuottaja) mahdollistaa tiedon uudelleenkäytön ja jatkojalostamisen selkeillä käyttöehdoilla (Creative Commons tai vastaava)	*	*	*

Tiedon rakenne on kuvattu selkeästi ja metatietostandardien mukaisesti.	*	- (aineistosta suuri osa tekstimuotoista)	*
---	---	---	---

Taulukko 1: Avoimen tiedon julkaisemiskriteerien toteutus Suomessa, Tanskassa ja Virossa

6 AVOIMEN TIEDON SAATAVUUS – ANALYYSI

6.1.1 Avoimen tiedon aineistojen analyysistä

Avoimen tiedon aineistojen analyysissä on tarkasteltu kansallisiin dataportaaleihin ja Euroopan dataportaaliin europeandataportal.eu sisältyviä tietoaineistoja. Dataportaaleissa on omina aihealueinaan kaksi tässä tarkasteltavaa teemaa, eli liikennetieto ja ympäristötieto. Eurooppalaisen dataportaalin aineistoa on myös verrattu kansallisten dataportaalien aineistoon, sillä kaikkien kansallisten portaalien aineistot eivät ole mukana eurooppalaisessa portaalissa. Aineistoja käsitellään luvussa 2 esitelyjen arviointikriteerien avulla (Berners-Leen kriteerit ja Hyvösen kriteerit).

6.1.2 Tanska

Ympäristötieto

Tanskassa ympäristötietoa koskevan avoimen tiedon seuranta on tehostettu viime vuosina. Tanskan valtiontilintarkastajien (Rigsrevisionen) raportissa on käsitelty avoimen tiedon tarjontaa ministeriökohtaisesti (Folketinget 2019). Ympäristöministeriön osalta raportista käy selville, että ympäristöasioita koskevia tietoaineistoja on tarjolla kaikkiaan 94 (Folketinget 2019, 62).

Haettaessa ympäristötietoa Tanskan avoimen tiedon portaalista opendata.dk, hakusanalla ”miljø” löytyi kaikkiaan 127 tietoaineistoa. Suurin osa näistä oli kahden yksittäisen kunnan tuottamia (Kööpenhamina ja Vejlen) ja avoimena datana on mm. tiedot parkkipaikoista, pyöräilyreiteistä, viheralueista ja julkisista rakennuksista. Tiedosta osa on karttatiedoissa käytettävissä WMS (Web Map Service) ja WFS (Web Feature Service)-muodoissa. Noin viidennes tietoaineistosta on avoimen tiedon jakelussa laajalti käytettävässä JSON-muodossa ja osa pelkinä pdf-tiedostoina.

Liikennetieto

Haettaessa liikennetietoa Tanskan avoimen tiedon portaalista, hakusanalla ”transport” tai ”trafik” löytyi kaikkiaan 216 tietoaineistoa. Aineistot koostuvat mm. kuntien reaaliaikaisesta liikennetiedosta, pysäköintilupa-alueita koskevista tiedoista, polkupyörien pumppauspistetiedoista, liikenneonnettomuustiedoista, jne. Tietoaineistosta yli puolessa on käytössä avoimen tiedon jakelun JSON-muoto tai GeoJSON-muoto.

6.1.3 Viro

Ympäristötieto

Haettaessa ympäristötietoa Viron avoimen tiedon portaalista opendata.riik.ee, hakusanalla ”keskkond” löytyi kaikkiaan 23 tietoaaineistoa. Nämä aineistot sisältävät mm. säätietoja, tietoa Viron vesihuollosta sekä tietoja Tallinnan ja Tarton kaupunkien palveluista (yleiset liikuntapaikat, wc-tilat, julkinen liikenne). Tietoaaineistot ovat saatavilla vain html-muodossa verkkosivuilla (veebileht) tai xml-muodossa (kaksi tietoaaineistoa). Avoimen tiedon JSON-muodossa on saatavilla vain teknisen valvontaviranomaisten tietojärjestelmä.

Liikennetieto

Haettaessa liikennetietoa Viron avoimen tiedon portaalista, hakusanalla ”transport” löytyi kaikkiaan 12 tietoaaineistoa. Nämä tietoaaineistot sisältävät mm. kansallisen osoitetietokannan, Tallinnan ja Tarton palveluhakemiston sekä elinkeinolupia koskevia tietoja.

Tietoaaineistot ovat saatavilla vain html-muodossa verkkosivuilla (veebileht) tai xml-muodossa (elinkeinotiedot). Viron Smart Road-aineisto on saatavilla erillisen API-rajapintasovelluksen avulla. Tarton julkisen liikenteen tiedot ovat saatavissa liikennejärjestelmien käyttämässä GTFS-muodossa (General Transit Feed Specification). Muita avoimen tiedon standardimuotoja ei ole käytössä.

6.1.4 Suomi

Ympäristötieto

Haettaessa ympäristötietoa Suomen avoimen tiedon portaalista, hakusanalla ”ympäristö” löytyi kaikkiaan 469 tietoaaineistoa. Nämä aineistot sisältävät tietoa mm. ympäristölupakohteista, ympäristöraporteista, liikennemelusta, ilmanlaadusta, luontokohteista, ulkoilureiteistä, jne.

Näistä tietoaaineistoista noin 25% on pelkkiä excel-tiedostoja. Avoimen tiedon jakelumuodoista on käytössä karttatiedoissa WFS ja WMS sekä JSON tai JSON/REST-muoto.

Liikennetieto

Haettaessa liikennetietoa Suomen avoimen tiedon portaalista, hakusanalla ”liikenne” löytyi kaikkiaan 354 tietoaaineistoa. Nämä aineistot sisältävät tietoa mm. liikennepaikoista, liikenneväylistä, kaupunkipyöristä, liikennetilastoista, liikennevaloristeyksistä, liikenneonnettomuuksista, jne.

Näistä tietoaaineistoista noin 15% on pelkkiä excel-tiedostoja. Avoimen tiedon jakelumuotoista on käytössä karttatietojen WFS ja WMS sekä JSON/REST-muoto.

6.1.5 Avoimen tiedon saatavuus - vertailua

Tanskan, Viron ja Suomen avoimen tiedon saatavuutta voidaan analysoida käyttämällä Berners-Leen viiden tähden kriteereitä, jotka ovat yleisimmin käytetty väline avoimen tiedon laadun arvioinnissa. Viiden tähden kriteereihin on yleensä liitetty käytössä olevat tiedostomuodot. Yhden tähden tasolla käytetään yleisimpinä tiedostomuotoina pelkkiä tekstitiedostoja (txt-, html- tai pdf-muodossa). Kahden tähden tasolla on käytössä esimerkiksi excel-tiedostoja. Kolmen tähden tasolla käytetään järjestelmäriippumattomia muotoja kuten csv tai xml. Neljän tähden tasolla ovat käytössä semanttisen verkon standardeihin perustuvia tiedostomuotoja ja kyselykieliä, kuten RDF, SPARQL, JSON tai erillisovelluksia (API) (Frosterus et. al. 2011, 96-97; Hyvönen 2018, 115-116).

Avoimen tiedon laatuksikriteerijä käyttäen voidaan tilanne havainnollistaa seuraavasti:

Kriteeri	Tanska		Viro		Suomi	
	Liikenne 216	Ympäristö 127	Liikenne 12	Ympäristö 23	Liikenne 354	Ympäristö 469
*saatavilla verkossa (missä tahansa tiedostomuodossa, txt,html, pdf, jpg) avoimella lisenssillä	linkki 14 html 47 pdf 24 docx 3	linkki 13 html 17 pdf	linkki html pdf docx		pdf 12 html	pdf 29 html 14 zip 7
** saatavilla koneluettavassa rakenteisessa muodossa (esimerkiksi excelinä eikä vain kuvatiedostona)	excel 63 docx 3 ods xyz	excel 21			excel 35	excel 44 docx 4

*** kuten ** mutta järjestel- märiippumatto- massa formaa- tissa (esim. CSVnä eikä excelinä)	csv 109 kml 10 wfs 58 shp 66 wms 53 wmts 43	csv 45 shp 43 wfs 29 wms 23 kml 20 xyz 14 wmts 12 kmz 7	api 1 gtfs 1 xml 9	xml 8	csv 45 wms 37 wfs 30 shp 24 kml 17 geoser- vice 13 xml 11 graphql 9 tab 6	csv 27 wms 83 wfs 60 xml 48 shp 32 pxweb 31 pc-axis 30 tab 22 kml 9
**** lisäksi: käytetty W3Cn avoimia stan- dardeja (RDF ja SPARQL) iden- tifioidaan tieto- ja, niin että jul- kaistuu tietoon voidaan viitata	GeoJSON 136	GeoJSON 57 JSON 11	JSON 1	JSON 1	GeoJSON 16 JSON 17	JSON 13 JSON/ REST 7

Taulukko 2: Avoin tiedon tiedostomuodot Suomessa, Tanskassa ja Virossa liikenne- ja ympäristöaineistoissa

Selvityksen perusteella erot yhtäältä Tanskan ja Suomen ja toisaalta Viron välillä ovat selvät. Tanskassa ja Suomessa on kohtalainen määrä avoimen tiedon tietoaaineistoista tarjolla muodossa, joka mahdollistaa tiedon monipuolisen hyödyntämisen ja liittämisen avoimen tiedon sovelluksiin (geotietoon liittyvä GeoJSON, JSON ja JSON/REST). Tanskassa jopa puolet tietoaaineistoista on saatavilla JSON-muodossa ja Suomessakin noin 5% tietoaaineistoista on tarjolla JSON-muodossa. Virossa avoimesta tiedosta suurin osa on tarjolla pelkästään html- tai xml-muodossa eikä niissä hyödynnetä avoimen tiedon metatietostandardeja.

7 TUTKIMUSTULOKSET JA PÄÄTELMÄT

Avoimen tiedon strategioita ja käytännön jakelua koskevan tutkimuksen keskeiset tulokset koskevat ensinnäkin eroja kansallisten avoimen tiedon strategioiden välillä ja toiseksi eroja avoimen tiedon jakelussa ja etenkin yhdistetyn avoimen tiedon saatavuudessa. Kansallisen avoimen tiedon strategioista on Tanskan strategiassa viety avointa tietoa koskeva analyysi kaikkein pisimmälle (Digitaliseringsstrategi 2016), vaikkakin toimenpidesuositukset ovat suppeita. Tanskan strategiassa on selkeästi korostettu Suomea tai Viroa enemmän yhteiskunnan perustietovarantojen (Grunddata) merkitystä ja hyödyntämistarpeita. Virossa on avoimen tiedon julkaiseminen nähty melko pienenä osana yhteiskunnan digitalisaatiota. Virossa on enemmän korostettu palvelujen nopeaa digitalisointia, oletusarvoista digitalisaatiota eli *digital by default* -periaatetta sekä kansalaisten digiosaamisen kehittämistä. Suomessa avoimen tiedon jakelu on ollut osa yleisempää tietovarantojen avaamista. EU:n yleinen tietosuoja-asetus on vaikuttanut henkilötietojen suojaamisen laajempaan käsittelyyn. Viron linjauksissa on ollut esillä tiedon anonymisointi avoimen tiedon jakelun edellytyksenä (Eesti Avaandmete 2016). Suomessa on jonkin verran painotettu yksityisyyden suojaa, mutta tiedon anonymisointitarpeisiin on kiinnitetty huomiota vasta avoimen tiedon tuottamisen edellytyksiä koskevassa selvityksessä (Koski et al. 2017, s.59). Tanskassa ja Suomessa on korostettu avoimen tiedon käyttöehtojen ja rajapintojen merkitystä, mutta Virossa käyttöehtoja on käsitelty lähinnä tiedon yhdenvertaisen jakelun kannalta.

Avoimen tiedon strategioissa ovat keskeiset erot avoimen tiedon tarjontaa koskevissa asioissa sekä avoimen tiedon standardien ja rajapintojen hyödyntämisessä. Tanskassa ja Suomessa avoimen tiedon tarjonnan linjaukset on jo keskeisiltä osin tehty, mutta Virossa linjaukset ovat alkaneet täsmenntyä vasta avoimen tiedon portaalin avaamisesta vuodesta 2018 lähtien. Erot avoimen tiedon hallinnointia koskevissa asioissa ovat myös merkittäviä. Virossa valtion rooli on keskeisempi kuin Suomessa tai Tanskassa. Tanskassa avoimen tiedon portaalista vastaavat pääosin kunnat. Myös avoimen tiedon jakelun standardien käytössä ja avoimen tiedon laajuudessa erot Tanskan, Suomen ja Viron välillä ovat merkittäviä. Tanskassa ja Suomessa on yleisessä käytössä kehittyneitä avoimen tiedon tiedostomuotoja (mm. JSON/REST), mutta Virossa avoin tieto julkaistaan pääosin verkkosivuilla html-muodossa. Myös julkaistujen tietoaineistojen määrissä on selviä eroja. Tanskassa ja Suomessa etenkin kaupungit ovat julkaisseet avoimen tiedon tietoaineistoja todella monipuolisesti. Virossa vain suurimmat kaupungit (Tallinna ja Tartu) ovat lähteneet liikkeelle avoimen tiedon julkaisemisessa. Virossa avoimen tiedon laaja tarjonta on vasta käynnistymässä, kun Tanskassa ja Suomessa avoimen tiedon tuottamisessa ja jakelussa on jo päästy vakiintuneisiin toimintatapoihin.

LÄHTEET

- Attard, J., Olandi, F. Auer, S. 2016. Value Creation on Open Government Data. IEEE Explore Digital Library. <http://ieeexplore.ieee.org/document/7427509/> <Käytetty 4.4.2020 >
- Bengtsson, M. 2016. How to plan and perform a qualitative study using content analysis. Nursing-Plus Open Vol. 2, 2016, pp. 8-14.
- Berners-Lee, T. 2012. Five-star deployment scheme for Open Data. <http://5stardata.info/en/> <Käytetty 8.4.2020 >
- Danish Ministry of Finance. 2019. National Strategy of Artificial Intelligence. Copenhagen.
- Eesti Avaandmete Portaali. 2016. Avaandmete loomise ja avaldamise juhend (Eesti Avaandmete 2016)
- Digitaliseringsstyrelsen. 2016. Et Stærkere og Mere Trygt Digitalt Samfund. Den fællesoffentlige digitaliseringsstrategi 2016-2020. Digitaliseringsstyrelsen: København (Digitaliseringsstrategi 2016) <https://digst.dk/strategier/digitaliseringsstrategien/> <Käytetty 7.4.2020 >
- Elo S., Kääriäinen M., Kanste O., Pölkki T., Utriainen T., Kyngäs H. 2014. Qualitative Content Analysis: A Focus on Trustworthiness. SAGE Open, January-March 2014: 1 –10.
- Estonian Ministry of Economic Affairs and Communications. 2016. Digital Agenda 2020 for Estonia. Tallinn. (Estonian Digital Agenda 2020)
- Euroopan komissio. 2011. Euroopan komission tiedonanto: Avoin data - Innovoinnan, kasvun ja läpinäkyvän hallinnon moottori. KOM(2011) 882 lopullinen. Bryssel 2011
- Euroopan komissio. 2020. Komission tiedonanto Euroopan parlamentille, neuvostolle, Euroopan talous- ja sosiaalikomitealle ja alueiden komitealle. Euroopan datastrategia. COM/2020/66 final
- Euroopan parlamentin ja neuvoston direktiivi 2003/98/EY, 17.11.2003 julkisen sektorin hallussa olevien tietojen uudelleenkäytöstä (EYVL 345, 31.12.2003, 90) (PSI-direktiivi)
- Euroopan parlamentin ja neuvoston direktiivi (EU) 2019/1024, 20.6.2019, avoimesta datasta ja julkisen sektorin hallussa olevien tietojen uudelleenkäytöstä (EUVL L 172, 26.6.2019, 56-83) (Avoin data –direktiivi)

European Data Portal 2020. The Economic Impact of Open Data. Opportunities for Value Creation in Europe. <https://www.europeandataportal.eu/sites/default/files/the-economic-impact-of-open-data.pdf> < Käytetty 29.03.2020 >

Folketinget Rigsrevisionen. 2019. Beretning 12/2018 om åbne data. Folketinget: København (Folketinget 2019)

Frosterus, M., Hyvönen, E., Laitio, J. 2011. Creating and Publishing Semantic Metadata about Linked and Open Datasets , teoksessa: Wood, D. (toim.) Linking Government Data, Cham: Springer, ss. 95-112.

Heinemann, L. 2010. Kirjastodataa raakana? Näkökulmia bibliografisen datan avaamiseen. Tietolinja 2/2010.

Horst, N.L., 2014. The Basic Data Programme – A Danish Infrastructure Model for Public Data.

Huijboom, N., Van den Broek, T. 2011. Open data: an international comparison of strategies. European Journal of E-Practice. No. 12, March/April 2011.

Hyvönen E., Tuominen J., Alonen M., Mäkelä E., 2014. Linked Data Finland: A 7-star Model and Platform for Publishing and Re-using Linked Datasets. Lecture Notes in Computer Science, vol . 8798. Cham: Springer.

Hyvönen E., 2018. Semanttinen Web. Linkitetyn avoimen datan käsikirja. Helsinki: Gaudeamus.

Implementation of the Public Sector Information Directive. European Commission. <https://ec.europa.eu/digital-single-market/en/implementation-public-sector-information-directive> (Implementation 2017). <Käytetty 03.04.2020>

Jaatinen, T. 2014. Julkisen sektorin avoin data ja henkilötietojen suoja. Pro gradu –tutkielma, Helsingin yliopisto, Helsinki 2014.

Kansalliskirjaston tiedon avoimen saatavuuden periaatteet. 2011. <http://www.doria.fi/handle/10024/88839> (Kansalliskirjaston periaatteet 2011).

Kassen, M. 2018. Adopting and managing open data: Stakeholder perspectives, challenges and policy recommendations. Aslib Journal of Information Management, vol. 70, nro 5, 518-537. (Kassen 2018a)

Kassen, M. 2018. Open data and e-government – related or competing ecosystems: a paradox of open government and promise of civic engagement in Estonia. *Information Technology for Development*, 25 (3), 552–578. (Kassen 2018b)

Kassen, M. 2019. *Open Data Politics: A Case Study on Estonia and Kazakhstan*. Bern:Springer

Knuuti, M. 2015. *Avoin data ja PSI-direktiivi julkisuuden toteuttajina*. Pro gradu tutkielma, Lapin yliopisto 2015.

Koski, H., Honkanen, M., Luukkonen J., Pajarinen, M., Ropponen T. *Avoimen tiedon hyödyntäminen ja vaikuttavuus*. Valtioneuvoston selvitys- ja tutkimustoiminta. Selvitys- ja tutkimustoiminnan julkaisusarja 40/2017. Helsinki 2017

Krippendorff, K., 2012. *Content analysis: an introduction to methodology*. New York: Sage.

Liikenteen ja viestinnän avoin tieto. Työryhmän raportti, LVM Julkaisuja 10 / 2013. <https://www.lvm.fi/-/liikenteen-ja-viestinnan-avoin-tieto-tyoryhman-raportti-811423> < Käytetty 29.03.2020 >

Lohtander H., 2013. *Avoin data ja semanttinen verkko – yhdessä kohti älykkäämpää internetiä* . Pro gradu tutkielma. Tampereen yliopisto, informaatiotieteiden yksikkö.

Lubarsky, B. 2017. Re-Identification of “Anonymized” Data. *Georgetown Law Technology Review* 202 , April 2017.

Margetts, H, Naumann A. 2017. *Government as a Platform. What can Estonia show the world?* Oxford Internet Institute. <https://www.politics.ox.ac.uk/materials/publications/16061/government-as-a-platform.pdf>

Mayring, P. 2014. *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. Klagenfurt. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-395173>

OECD. 2015. *Public Governance Reviews. Estonia and Finland - Fostering Strategic Capacity across Governments and Digital Services across Borders*. Paris: OECD.

Open Data Denmark. <http://www.opendata.dk> <Käytetty 2.4.2020 >

Open Data Maturity in Europe 2019. *Insights into the European state of play*. European Commission, Cap Gemini Consulting. Brussels: European Commission.

Open Government Data. 2007. <https://opengovdata.org/> <Käytetty 02.04.2020>

Poikola A., Kola, P., Hintikka, K. 2010. Julkinen data. Johdatus tietovarantojen avaamiseen. Liikenne- ja viestintäministeriö, Helsinki.

PriceWaterhouseCoopers. 2019. Open data and e-government good practices for fostering environmental information sharing and dissemination. Copenhagen: European Environment Agency.

Røyen, J.K., 2014, The Danish Basic Data Programme and the governance perspective. Copenhagen: Digitaliseringsstyrelsen,

Simperl, E., O'Hara, K, Gomer R. 2016. Open Data and Privacy. European Data Portal (https://www.europeandataportal.eu/sites/default/files/open_data_and_privacy_v1_final_clean.pdf)

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

Valtioneuvosto. 2011. Valtioneuvoston periaatepäätös 03.03.2011 julkisen sektorin digitaalisten tietoaineistojen saatavuuden parantamisesta ja uudelleenkäytön edistämisestä. (VN-periaatepäätös 2011) <https://www.lvm.fi/-/julkisen-tietoaineiston-saatavuudesta-periaatepaatos-784401>

Valtiovarainministeriö. 2013. Julkisen hallinnon tietovarantojen avaamisen periaatteet. Valtiovarainministeriön muistio VM041:00/2013.

Valtiovarainministeriö. 2016. Avoimen datan opas. <https://www.avoindata.fi/fi/opas> (VM Avoimen datan opas)

Valtiovarainministeriö. 2015. Avoimesta datasta innovatiiviseen tiedon hyödyntämiseen. Avoimen tiedon ohjelman 2013–2015 loppuraportti. Valtiovarainministeriön raportti 31/2015. Helsinki (VM Avoimen tiedon ohjelman loppuraportti 2015).

Valtiovarainministeriö. 2015. Avoimen tiedon vaikuttavuus. Valtiovarainministeriön julkaisuja 15a/2015. Helsinki. (VM Avoimen tiedon vaikuttavuus)

Wood, D. (toim.). Linking Government Data. 2011. Cham: Springer.