

Heli Koski, Mika Honkanen, Juha Luukkonen,
Mika Pajarinen, Teemu Ropponen

Avoimen datan hyödyntäminen ja vaikuttavuus

Maaliskuu 2017

Selvitys- ja
tutkimustoiminnan
julkaisusarja 40/2017

KUVAILULEHTI

Julkaisija ja julkaisuaika	Valtioneuvoston kanslia, 16.3.2017
Tekijät	Heli Koski, Mika Honkanen, Juha Luukkonen, Mika Pajarinen, Teemu Ropponen
Julkaisun nimi	Avoimen datan hyödyntäminen ja vaikuttavuus
Julkaisusarjan nimi ja numero	Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 40/2017
Asiasanat	Avoin data, datatalous, tietojohdaminen, vaikuttavuus

Julkaisun osat/ muut tuotetut versiot

Julkaisuaika	2017	Sivuja 89	Kieli Suomi
---------------------	------	------------------	--------------------

Tiivistelmä

Tässä raportissa kartoitetaan keskeisten perustietovarantojen avoimuutta, selvitetään mahdollisuuksia avoimen datan taloudellisten vaikutusten arviointiin sekä valotetaan avoimen datan käytön ja yritysten innovaatioiden tuottamisen sekä taloudellisen menestyksen välistä yhteyttä. Lisäksi raportissa esitetään konkreettisia ehdotuksia, joiden avulla perustietovarantojen vaikuttavuutta yhteiskunnassa voidaan lisätä.

Tutkimuksemme osoittaa, että avointa dataa ja massadataa innovaatiotoiminnassaan hyödyntävät yritykset tekevät markkinoiden kannalta uusia tuoteinnovaatioita suhteellisesti huomattavasti useammin kuin yritykset, jotka eivät hyödynnä dataa. Dataa innovaatioiden kehittämisessä käyttäneiden informaatio- ja viestintätoimialan yritysten liikevaihto kasvoi vuosina 2012–2014 keskimäärin yli 17 prosenttia enemmän kuin saman toimialan dataa innovaatioissa hyödyntämättömien yritysten liikevaihto. Yksittäisistä tietotyypeistä vain liikennetietojen hyödyntäminen innovaatiotoiminnassa oli yhteydessä yritysten liikevaihdon kasvuun.

Datatalouden mekanismien ymmärtämiseksi tarvittaisiin tutkimustietoa yritysten datan käytöstä ja taloudellisista vaikutuksista laajemmin. Tätä varten tulisi kehittää datan käyttöä kuvastavaa tilastoaineistojen keräämistä. Avoimen tiedon hyödyntämisedellytysten parantamiseksi tarvitaan hallinnonalat ylittäviä parhaita käytäntöjä tiedonhallintaan ja datan avaamiseen liittyen. Eri virastojen ja hallinnonalojen tiedonhallinta on verrattain epäyhtenäistä. Avoimen datan jakelu on pirstaleista, jonka takia sen löydettävyyttä on datan hyödyntäjille haasteellista. Hyviä työkaluja tiedonhallinnan jäsentämiseen, kuten esimerkiksi tietotilinpäätös on kehitetty, mutta niitä ei ole vielä otettu laajasti käyttöön. Henkilötietoon liittyvät käytännöt kaipaavat tarkennusta. Datatietojen hyödyntämistä voidaan lisätä mm. lisäämällä vuorovaikutusta tietovarantojen hallinnoijien ja hyödyntäjien kanssa.

Tämä julkaisu on toteutettu osana valtioneuvoston vuoden 2016 selvitys- ja tutkimussuunnitelman toimeenpanoa (tietokayttoon.fi).

Julkaisun sisällöstä vastaavat tiedon tuottajat, eikä tekstisisältö välttämättä edusta valtioneuvoston näkemystä.

PRESENTATIONSBLAD

Utgivare & utgivningsdatum Statsrådets kansli, 16.3.2017

Författare Heli Koski, Mika Honkanen, Juha Luukkonen, Mika Pajarinen, Teemu Ropponen

Publikationens namn Användningen och effektivitet av öppna data

Publikationsseriens namn och nummer Publikationsserie för statsrådets utrednings- och forskningsverksamhet 40/2017

Nyckelord Öppna data, data ekonomi, data management, effektivitet

Publikationens delar /andra producerade versioner

Utgivningsdatum	2017	Sidantal	89	Språk	Finska
------------------------	------	-----------------	----	--------------	--------

Sammandrag

I föreliggande rapport kartläggs öppenheten i de viktigaste basregistren, möjligheterna till systematisk insamling av konsekvensbedömningsdata samt redogörs för sambanden mellan användning av öppna data och företags innovation och ekonomiska framgångar. I rapporten föreslås också konkreta åtgärder för att stärka basdatalagens roll i samhället.

Vår analys av materialet visar att företag som utnyttjar öppna data or stordata (massadata) i sin innovationsverksamhet tar fram varu- och/eller tjänsteinnovationer med marknadsbetydelse relativt sett betydligt oftare än företag som inte utnyttjar data. Rönen visar också att de företag inom informations- och kommunikationssektorn som hade använt data för framtagning av innovationer nådde en i genomsnitt över 17 procent större omsättningsökning åren 2012–2014 än företag inom samma sektor som inte hade utnyttjat data. Av de enskilda datatyperna var endast utnyttjande av kommunikationsdata för innovationsverksamhet kopplat till omsättningsökning.

För att vi ska kunna förstå de dataekonomiska mekanismerna behövs det mer forskningsrön om företags användning av data och dess ekonomiska konsekvenser överlag. Detta förutsätter mer utvecklad insamling av statistik om dataanvändning. För att förbättra möjligheterna för utnyttjande av öppna data behövs förvaltningsövergripande bästa praxis för informationshantering och öppnande av data. Samordningen i fråga om informationshanteringen på ämbetsverken och inom förvaltningarna är rätt så bristfällig. Distributionen av öppna data är splittrad och de som utnyttjar data har svårt att få fram data. Det finns bra verktyg för strukturerad informationshantering, som bland annat databokslutet, men dessa har inte tagits i bruk i någon större omfattning. De rutiner som anknyter till hantering av personuppgifter bör specificeras. Det är möjligt att öka utnyttjandet av data bland annat genom att öka interaktionen mellan dem som förvaltar och dem som använder datatillgångarna.

Den här publikation är en del i genomförandet av statsrådets utrednings- och forskningsplan för 2016 (tietokaytoon.fi/sv).

De som producerar informationen ansvarar för innehållet i publikationen. Textinnehållet återspeglar inte nödvändigtvis statsrådets ståndpunkt.

DESCRIPTION

Publisher and release date	Prime Minister's Office, 16.3.2017
Authors	Heli Koski, Mika Honkanen, Juha Luukkonen, Mika Pajarinen, Teemu Ropponen
Title of publication	Use and impacts of open data
Name of series and number of publication	Publications of the Government's analysis, assessment and research activities 40/2017
Keywords	Open data, data economy, data management, impacts

Other parts of publication/ other produced versions

Release date	2017	Pages	89	Language	Finnish
---------------------	------	--------------	----	-----------------	---------

Abstract

This report sheds light on the openness of major data resources maintained by the public administration and on means to assess the economic impacts of open data in Finland. It further studies the relationship between firms' use of open data and their innovation production and growth. Furthermore, this report proposes concrete means how to enhance the impact of open data in our society.

Our empirical analysis shows that firms using open data and big data in their innovation activities were clearly more likely to produce new-to-market innovations than those firms that didn't use open data in their innovation activities. Also, we found that those firms active in information and communications sector and using data for innovation witnessed, on average, 17 percent higher sales growth during the years 2012–2014 than those firms in the same sector that didn't use data in innovation. Traffic data appeared to be the only data type contributing to the sales growth.

More in-depth research on firms' data use and its economic impacts is essential to enhance our understanding of the mechanisms of data economy. For that, it is necessary to develop systematic collection of statistical data concerning data usage indicators.

Best practices on data management and open data publishing across administrative sectors are needed for improving conditions for the effective utilization of open data. Currently, the data management practices adopted by different organisations are somewhat incoherent. The delivery of open data is fragmented which makes data difficult to find. There are certain suitable tools such as data balance sheet available for data management but they are not yet widely adopted in Finland. Practices related to personal data require further evaluation and clarification. Utilization of data could be promoted by, for instance, increasing engagement between the data holders and data users.

This publication is part of the implementation of the Government Plan for Analysis, Assessment and Research for 2016 (tietokayttoon.fi/en).

The content is the responsibility of the producers of the information and does not necessarily represent the view of the Government.

SISÄLLYS

1	Johdanto	7
2	Rekisteritietojen tarjonta, avoimuus ja hyödyntäminen	9
	2.1 Rekisteritietojen tarjonta Suomessa	9
	2.2 Perustietovarantojen avaaminen ja käyttö	12
3	Avoimen datan taloudellinen vaikuttavuus	18
	3.1 Minkälaisia mittareita tarvitsemme, onko meillä dataa?	18
	3.2 Miten taloudellista vaikuttavuutta voidaan tutkia?	21
	3.3 Aineistoanalyysin tulokset	23
	3.3.1 Avoimesta datasta innovaatioita?	23
	3.3.2 Luoko datan käyttö innovaatioissa kasvua?	27
4	Avoimen datan hyödyntäminen	32
	4.1 Tietojen kerääminen	32
	4.2 Johtaminen ja ohjaus	33
	4.2.1 Valtioneuvoston digitalisaation tiedonhallinnan tavoitteet	33
	4.2.2 Tulosohejaus ohjaamaan tietovarantojen avaamista ja hyödyntämistä	35
	4.2.3 Tietotilinpäätös tiedonhallinnan ja avoimen datan apuvälineenä	36
	4.2.4 Organisaatioiden aineisto- tai datapolitiikat	36
	4.2.5 Kooste avoimen tiedon ohjauskeinoista	38
	4.3 Avoimen datan kustannukset ja lisäarvo	38
	4.3.1 Datan avaamisen kustannukset	38
	4.3.2 Taloudellinen resursointi datan avaamiseen ja ylläpitoon	40
	4.3.3 Datan avaamisen taloudelliset vaikutukset	41
	4.4 Tehokasta datan hyödyntämistä tukevat toimintatavat	43
	4.4.1 Datan avaamisen priorisointi kysyntä- ja vaikuttavuuslähtöisesti	43
	4.4.2 Datan avaajien ja hyödyntäjien välinen vuorovaikutus	45
	4.4.3 Avoimuudella irti toimittajalukoista	46
	4.4.4 Osaamisen kehittäminen ja parhaiden käytäntöjen jakaminen	47
	4.4.5 Tehostuneet digitaaliset prosessit	47
	4.5 Avoimen datan lainsäädäntö	48
	4.5.1 Avoimet lisenssit tiedon hyödyntämisen mahdollistajana	49
	4.5.2 Haasteena vaihtelevat lain tulkinnat	49
	4.5.3 Henkilötiedon ja avoimen datan välinen suhde kaipaa selvennyksiä	50

4.6	Tietoinfrastruktuuuri ja teknologia.	51
4.6.1	Avoindata.fi ja dataportaalit.	52
4.6.2	Katselu- ja latauspalvelut ja ohjelmointirajapinnat	55
4.6.3	Avoimen datan laadun parantaminen	56
4.6.4	Dataspektri kuvaa tiedon avoimuuden koko skaalaa	58
4.6.5	Henkilötiedon hyödyntäminen.	59
5	Toimenpiteitä avoimen datan vaikuttavuuden lisäämiseksi	63
6	Päätelmät	69
	Lähdeluettelo	71
Liite 1	Keskeiset käsitteet.	75
Liite 2	Etl:n kyselylomake rekisterinpitäjille	80
Liite 3	Datan avaajien teemahaastattelurunko	83
Liite 4	Lähteet – Rekisteritietojen tarjonta Suomessa	86

1. JOHDANTO

Suomessa julkishallinnon tietovarantojen avaaminen¹ on linjattu tavoitteeksi useissa yhteyksissä². Tietovarantojen avaamista vauhditettiin vuosina 2013–2015 valtiovarainministeriön asettaman avoimen tiedon ohjelman toimenpitein ja informaatio- ja talousohjauksen keinoin. Lähtötavoitteena on ollut avata vuosikymmenen loppuun mennessä kaikki merkittävät julkishallinnon julkiset tietovarannot ottaen huomioon lainsäädännön rajoitukset. Viranomaisten tietoaineistoja on avattu monipuolisesti ja laajasti: esimerkiksi paikkatietoaineistoja, sää- ja liikennedataa, ajoneuvotietoja, yhteisöverotietoja, kulttuuriaineistoja ja lainsäädännöllistä tietoa. Esimerkiksi maanmittauslaitoksen digitaaliset maastotiedot avattiin toukokuussa 2012. Vuotta myöhemmin Ilmatieteen laitos avasi säätiedot avoimeksi dataksi. Vuonna 2014 avattiin Avoindata.fi, avoimen tiedon ja yhteentoimivuuden palvelu, jonka tarkoituksena on kerätä tiedot kaikista avoimena datana tarjottavista aineistoista yhteen palveluun. (Kauhanen-Simanainen & Suurhasko, 2015, s. 1–15).

Raportoidun hankkeen tavoitteina oli i) kartoittaa missä määrin perustietovarannot on avattu, mitä ovat tärkeimmät hyödynnettävät tiedot ja mitä esteitä (pois lukien lakisääteiset esteet) tietovarantojen avaamiselle on edelleen olemassa, ii) selvittää avoimen datan taloudellisen vaikuttavuuden arviointiin tarvittavien mittareiden saatavuutta sekä mahdollisuutta tarvittavien tietojen systemaattiseen keräämiseen, iii) tutkia kehittyneillä tilastollisilla menetelmillä avoimen datan käytön ja yritysten taloudellisen menestyksen välistä yhteyttä ja iv) pohtia ja tuottaa konkreettisia ehdotuksia, joiden avulla perustietovarantojen vaikuttavuutta yhteiskunnassa voidaan lisätä (esim. politiikkatoimet, uudet tai muuttuneet toimintatavat tai uudet tekniset muutokset). Tässä raportissa keskitytään avoimen datan taloudellisiin vaikutuksiin.

Julkisen hallinnon tietohallinnon neuvottelukunnan perustietovarantojaoston määritelmän mukaisesti perusrekisterit pitävät sisällään ajantasaista ja tarkkaa tietoa yhteiskunnan perusyksiköistä ja perusyksiköiden välisistä suhteista. Näitä perusyksiköitä ovat luonnolliset henkilöt, oikeushenkilöt, kiinteistöt ja rakennukset. Perusrekisterien pitämisestä säädetään laissa ja Suomessa perusrekistereitä ovat väestötietojärjestelmä (VTJ), kiinteistötietojärjestelmä (KTJ) sekä yritys- ja yhteisötietojärjestelmät. Perusrekisterit voivat pitää sisällään myös paikkatietoa sekä historiallisia tietoja perusyksiköistä. Perustietovarannoiksi voidaan laskea perusrekisterien lisäksi erilaisia muita keskeisiä kansallisia tietovarantoja. Tätä tutkimusta varten projektin ohjausryhmä valitsi tarkasteltavaksi edustavan joukon perusrekistereitä ja keskeisiä perustietovarantoja.

Luku 2 käsittelee tietojen avoimuutta valittujen perustietovarantojen osalta. Tietojen avoimuutta ja tarjontaa kartoitettiin ensinnäkin saatavilla olevien julkisten tietojen perusteella. Toiseksi perustietovarantoja hallinnoiville organisaatioille tehtiin syksyllä 2016 kysely, jossa kartoitettiin sitä, missä määrin perustietovarannot on avattu, mitä ovat tärkeimmät hyödynnettävät tiedot ja mitä esteitä (pois lukien lakisääteiset esteet) tietovarantojen avaamiselle on edelleen olemassa.

¹ Julkinen avoin data on julkisen organisaation tuottama tai hallinnoima tieto(varanto), jonka on 1) konekielisessä muodossa, 2) maksutta kenen tahansa käytettävissä ja 3) käytettävissä sekä yksityisiin että kaupallisiin tarkoituksiin. Julkisen avoimen datan voi nähdä tietoyhteiskunnan pääomana, jonka arvo kasvaa sitä jaettaessa ja jalostettaessa.

² Valtioneuvoston periaatepäätös 3.3.2011 julkisen sektorin digitaalisten tietoaineistojen saatavuuden parantamisesta ja Valtioneuvoston periaatepäätös datan hyödyntämisestä liiketoiminnassa 19.5.2016, pääministeri Kataisen, pääministeri Alexander Stubbin ja pääministeri Juha Sipilän hallitusohjelmien kirjaukset, Avoimen hallinnon hanke (Open Government Partnership) vuodesta 2013 lähtien, ministerityöryhmän hyväksymät digitalisoinnin periaatteet 2.2.2016.

Luku 3 keskittyy avoimen datan taloudellisen vaikuttavuuden mittaamiseen. Kosken (2015) esitutkimusraportissa esitetyn alustavan vaikuttavuusmittariston kehittämistä jatkettiin selvittämällä, miltä osin tietoa vaikuttavuuden arviointiin tarvittavista muuttujista on saatavilla sekä mahdollisuutta tarvittavien tietojen systemaattiseen keräämiseen. Luvussa esitetään myös toimenpiteitä, joita avoimen datan yhteiskunnallis-taloudellisen vaikuttavuuden pitkäjänteinen tutkiminen ja mittareiden kerääminen edellyttäisi. Luvussa 3 keskustellaan myös menetelmistä, joita voidaan käyttää avoimen datan taloudellisen vaikuttavuuden mittaamiseen. Luku 3 raportoi kahdella suomalaisella yritysaineistolla tehdyn empiirisen tutkimuksen tulokset koskien datan hyödyntämistä innovaatio toiminnassa ja sen yhteyttä yritysten innovaatioiden kehittämiseen ja kasvuun.

Luvussa 4 ja 5 esitetään ehdotuksia, joiden avulla perustietovarantojen vaikuttavuutta yhteiskunnassa voidaan lisätä. Näitä ovat esimerkiksi politiikkatoimet, uudet tai muuttuneet toimintatavat tai tekniset muutokset. Luvussa 4 pyritään myös identifioimaan ”korkean lisäarvon” tietovarantoja, joiden avaaminen olisi erityisen merkityksellistä. Esitetyjä toimenpide-ehtotuksia on kehitetty huomioiden esimerkiksi poliittiset, taloudelliset, toimintakulttuuriin liittyvät, teknologiset, oikeudelliset ja muut näkökulmat. Aineistona on käytetty muun muassa perustietovarantoja hallinnoivissa organisaatioissa tehtyjä haastatteluja ja verkkosivujen kautta kerättyä tietoa sekä työpajoissa tuotettuja näkemyksiä ja muuta materiaalia.

Luvussa 6 esitetään keskeiset johtopäätökset ja jatkotutkimusmahdollisuudet.

2. REKISTERITIETOJEN TARJONTA, AVOIMUUS JA HYÖDYNTÄMINEN

Heli Koski – Juha Luukkonen

Syksyllä 2016 tehdyn kyselytutkimuksen vastaukset osoittavat, että datan avaaminen on varsin pitkällä keskeisiä perustietovarantoja hallinnoivissa organisaatioissa. Vastaajaorganisaatioiden lainsäädännön puitteissa avattavissa olevista tietotuotteista oli avattu syksyyn 2016 mennessä keskimäärin 70 prosenttia. Vastanneiden organisaatioiden tietotuotteet olivat suurelta osin maksuttomia. Kustannussyyt ovat edelleen merkittävin este vielä avaamattomien aineistojen maksuttomalle tarjonnalle; datan avaamisen kustannusten kattaminen budjettirahoitteisesti lisäisi avoimen datan tarjontaa nykyisestä. Tietotuotteiden myynnistä saadut tulot vuonna 2015 poikkesivat verrattain vähän vuoden 2008 vastaavista tuloista, kun taas tietotuotteiden tuottamisen kustannukset olivat laskeneet huomattavasti. Tietotuotteiden suurin käyttäjäryhmä ovat yritykset: yritysten osuuden arvioitiin olevan keskimäärin liki puolet.

2.1 Rekisteritietojen tarjonta Suomessa

Perustietovarantojen avoimuutta ja luovutuskäytäntöjä keskeisten tietovarantoja hallinnoivien organisaatioiden osalta kuvataan taulukossa 2.1, joka on koostettu organisaatioiden nettisivuilta julkisesti saatavilla olevien tietojen mukaan kesällä 2016. Taulukossa on pyritty kuvailemaan keskeisten rekisteritietoja hallinnoivien organisaatioiden tarjoamien tietotuotteiden kirjon maksullisuutta, julkista saatavuutta sekä koneellista luettavuutta.

Tietotuotteita tarjotaan hyvin monimuotoisesti, ja tietovarantojen ylläpitäjät käyttävät tietotuotteidensa tarjonnassa lukuisia erilaisia rajapintoja ja tiedostomuotoja. Kaikki vertailussa mukana olevat organisaatiot tarjoavat jossakin muodossa koneluettavaa dataa, mutta sen luovutusmuoto saattaa vaihdella käyttäjäryhmittäin tai tarkemmin tietotuotteen aiotun käyttötarkoituksen mukaan. Variaatioita luovutettavien tietotuotteiden muodon välillä on käytännössä usean eri tiedostomuodon kattavista avoimista massapoinnista aina maksullisiin suppeampiin yhden havainnon kattaviin pdf- tai tekstiviestiotteisiin. Kaikkia tietotuotteita ei ollut tarjolla rakenteellisessa muodossa. Tietotuotteiden tarjoaminen on joissakin tapauksissa ulkoistettu yhteistyöyrityksille.

Kaikki tietovarantojen ylläpitäjät tarjoavat ainakin joitakin tietotuotteistaan avoimena datana. Yhteisenä nimittäjänä näissä ovat datan maksuttomuus, verkkosaatavuus, konekielinen luettavuus ja bulkkimuotoisuus. Avoimen datan tarjonnan laajuudessa on huomattavaa vaihtelua organisaatioiden välillä. Käytännössä kaikki otoksessa mukana olevat tietovarantoja hallinnoivat organisaatiot tarjoavat myös maksullisia ja/tai luvanvaraisia tietotuotteita. Tietoteknisesti edistyneempi käyttö niin erilaisten poimintojen kuin verkkokäytön osalta saattaa olla maksumuurin takana ja luvanvaraista, mutta esimerkiksi rekisteriotteisiin voi kaikilla halukkailla olla (maksullinen) pääsy. Tietotuotteiden luovutettavuus vaihtelee kohtalaisesti eri toimijoiden välillä ja etenkin maksullisen ja/tai luvanvaraisen datan osalta ei ole aina selvää, mitä tietoa tarkalleen on saatavilla, missä muodossa tieto on saatavilla ja mitä se maksaa. Tietotuotteiden tarjonnan läpinäkyvyydessä olisi tältä osin parannettavaa. Käytännön syitä monimutkaiseen vyyhteen, kuten organisaatioiden erilaisia tiedonhallinnan tiloja sekä lainsäädännön erilaisia tulkintoja, avataan tarkemmin raportin neljännessä luvussa, jossa käydään läpi teemahaastatteluita sekä työpajojen tuloksia.

Taulukko 2.1 Rekisteritietojen tarjonta Suomessa

Tietovastuullinen taho & rekisteri-/tietotuotteet	Sisältö	Verkkoosaavuus/julkisuus & koneellinen luettavuus	Ilmaisuus ja muu mahdollinen maksullinen data
VRK Väestötietojärjestelmä (VTJ): Henkilötiedot Rakennus ja huoneistotiedot	Perustiedot Suomen kansalaisista ja Suomessa vakinaisesti asuvista ulkomaalaisista. Järjestelmään rekisteröidään henkilöitä koskevat perustiedot, joita ovat muun muassa nimi, henkilötunnus, osoitetiedot, kotikunta, kansallisuus, perhesuhdetiedot sekä syntymä- ja kuolintieto. (1) Tietoja rakennuksista, rakennushankkeista ja huoneistoista. Muun muassa omistustietoja, kaavoitustietoja, valmistuspäivämääriä, rakennuslupia ja käyttötarkoituksia. (2)	Ei julkisesti saatavilla, vaatii tietoluvan VRK:ltä niin yritysten kuin julkishallinnon osalta. Tietolupahakemuksessa on ilmoitettava haettavien tietojen käyttötarkoitus ja hakemuksen yhteydessä on annettava selvitys tietojen käytöstä ja suojauksesta. (4) Ei julkisia tietoja bulkkimuotoisuudesta eikä koneellisesta luettavuudesta.	Maksullinen. Vakimuotoiset tietoluovutukset eivät maksa mitään viranomaisille tai tieteelliseen tutkimukseen ja lakisääteiseen tutkimukseen johtavaa koulutusta järjestäville oppilaitoksille, muut kuin vakimuotoiset luovutukset omakustannehinnoitella. (5) Ev.lut. sekä ort. kirkko saavat myös tietoluovutuksia omakustannus- tai alennetun omakustannusarvoisesti sekä yritykset tai yhteisöt, joilla on lakisääteinen tehtävä, saavat tietoluovutuksia omakustannusarvoisesti. Muille tahoille luovutettavien väestötietopalveluiden suhteen harjoitetaan liiketaloudellista hinnoittelua. (6) Osa tietopalveluista toteutetaan yhdessä yhteistyökumppanien kanssa, jotka huolehtivat muun muassa poimintojen teknisestä toteutuksesta ja rekisterien päivityksistä sekä tarjoavat suoramarkkinointipalveluja (3).
VRK Väestörekisterikeskuksen avoin data	VRK on julkaissut avoimena datana rakennusten osoitetiedot (sis. mm. pysyvät rakennustunnukset, tiedon käyttötarkoituksesta, koordinaattitiedot ja osoitteen ja äänestysalueen), Suomessa saatavilla olevien julkisten palvelujen hakemiston, suomalaisten nimiaineiston ja listauksen voimassa olevista ja lakaneista Suomen kunnista. (7)	Saatavilla julkisesti verkosta. <i>Tiedostomuotoja:</i> mm. xlsx, opt, pdf, riippuen datasta. (7)	Maksuton. (7)
MML Maastotietojärjestelmä (MTJ): Maastotietokanta Nimistörekisteri Laserkeilausaineisto, ilma- ja ortokuvat	Maastotietokanta: Koko Suomen kattava maastoa kuvaava aineisto. Tärkeimpiä kohderyhmiä ovat liikenneväyläverkosto, rakennukset ja rakenteet, nimistö, maankäyttö ja korkeussuhteet. (1) Nimistörekisteri: Suomen paikannimien ja kiinteistöjen ajantasainen tietovarasto, joka muodostuu paikannimirekisteristä ja karttanimirekisteristä. (3) Laserkeilausaineisto on maanpintaa ja maanpinnalla olevia kohteita kuvaava kolmiulotteinen aineisto. (4) Ilmakuva on oikaisematon rasterimuotoinen ilmakuva. (5) Ortokuvat ovat koko maan kattava ilmakuva-aineisto, joka vastaa geometrialtaan karttaa. (6)	Saatavilla julkisesti verkosta. (1) (3) (4) (5) (6) <i>Tiedostomuotoja:</i> Maastotietokanta: GML, MIF, shape, MAAGIS/XL (1) Nimistörekisteri: XML, GML (6) Laserkeilausaineisto, ilma- sekä ortokuvat: LAZ (4) TIFF, JPEG 2000 (vain orto) (5) (6)	Aineisto ja käsittely on maksutonta, kun aineisto toimitetaan ilmaisten aineistojen lataussivuston kautta itsepalveluna. Muussa tapauksessa erilliset toimitus- ja käsittelymaksut. (1) (3) (4) (5) (6) Virallisesti aineistojen luovutus on liiketaloudellisesti hinnoiteltu, mutta asetuksessa todetaan, että mikäli luovutuksen kustannukset ovat vähäiset, niin ne voidaan jättää perimättä. (2)
MML Kiinteistötietojärjestelmä (KTJ): Kiinteistörekisteri Lainhuuto- ja kiinnitysrekisteri Kiinteistöjen kauppahintarekisteri	Kiinteistötietoja ylläpitäviä rekistereitä. Kiinteistörekisteristä selviää esimerkiksi kiinteistön kiinteistötunnus, nimi, pinta-ala, sijainti, osuudet yhteisiin alueisiin ja käyttöoikeuksia ja -rajoituksia. (7) Lainhuuto- ja kiinnitysrekisteristä selviää kiinteistön omistaja sekä kiinteistöön kohdistuvat kiinnitykset ja erityiset oikeudet. (18) Kiinteistöjen kauppahintarekisteriin merkitään kaupanvahvistajien ilmoitusten perusteella tietoja luovutusten kohteista, kauppahinnoista sekä luovuttajista ja luovutuksen saajista. (20)	Ei julkisesti saatavilla ja vaatii käyttöluvan. Lupa voidaan myöntää ammatillisille käyttäjille ja viranomaisille (9) (14). Kiinteistötietojärjestelmän aineistopalvelu on tarkoitettu lähinnä kunnille ja kaupungeille (8) (18). Muille käyttäjille rajapintapalvelu Kiinteistötietopalvelu verkossa (9) (14). Muille asiakkaille pdf-tiedostot selainpohjaisen palvelun kautta (9) tai tulosteet (10) (19). Kiinteistöjen kauppahintarekisteri: PDF-tiedostot selainpohjaisen palvelun kautta (9) tai tulosteet (21). Tarvittaessa poimintoja voidaan tehdä tilauksesta ja säännöllisesti tarvittavia poimintoja voidaan tilata aineistopalveluna (21).	Kiinteistötietojärjestelmä: maksullinen ja käyttö edellyttää sopimusta. Käyttöluva voidaan myöntää mikäli tietoja tarvitaan yhdyskuntasuunnittelua, kiinteistönvälitystä, luoton myöntämistä taikka muuta näihin verrattavaa kiinteistöihin liittyvää tarkoitusta varten (13). Aineistopalvelu voidaan tilata kunnille, jotka pitävät kiinteistörekisteriä (11). Kiinteistötietojärjestelmän tiedot ovat viranomaisille sekä opetus- ja tutkimuskäyttöön maksuttomia (10) (12) (15). Yksityshenkilöt voivat tilata tietoja maksullisina otteina (10) (19). Sähköisten asiointijärjestelmien käyttö on hinnoiteltu omakustannusperusteisesti. (16) (17) Kiinteistöjen kauppahintarekisteri: maksullinen ja käyttö edellyttää sopimusta. Valtion viranomaisilta peritään maksu, joka vastaa aineiston irrottamisen ja luovuttamisen omakustannusarvoa (16). Yksityshenkilöt voivat tilata tietoja maksullisina otteina (21). Nimi- ja osoitetietoja voidaan luovuttaa tulosteena suoramarkkinointia ja markkinointitutkimuksia varten (22). Jos aineistoa käytetään muuten kuin viranomaistarkoituksiin, peritään maksu liiketaloudellisiin perustein. (16)
PRH Kaupparekisteri Säätiörekisteri Yhdistysrekisteri	Kaupparekisteri: yritystiedot sekä tiedot asunto-osakeyhtiöistä ja asumisoikeusyhdistyksistä. Yleensä yritysmuodosta rekisteröidään ainakin yrityksen nimi, kotipaikka, toimiala ja ketkä edustavat yritystä. Virre-tietopalvelusta saa tietoa yritysten voimassa olevista rekisteritiedoista ja yhteisöosäännöistä. Lisäksi historiatietoja ja lakanneiden yritysten tietoja tai tilinpäätöstietoja. (1)	Data on verkossa, mutta sen käyttö vaatii rekisteröitymisen ja se on maksuttomasti takana perustietoja lukuunottamatta. (3) (9) Kauppa- ja säätiörekisteri: Tiedostomuoto pdf erillisille yrityskohtaisille otteille (2). Myös erilliset poiminnat mahdollisia (3). Yhdistysrekisteri: PDF-muotoiset yritys- ja yhteisökohtaiset otteet. Osoitteet mahdollista poimia taulukoksi. (9)	Kaupparekisteri/Säätiörekisteri: Palvelu on maksullinen perustietoja lukuunottamatta (3). Erilaisia suurempia poimintoja voi myös suorittaa maksullisena ja näiden hinnat riippuvat poiminnan laajuudesta. (5) Vakimuotoiset tietoluovutukset kunnan ja valtion viranomaisille sekä tutkimukseen ja lakisääteiseen tutkimukseen johtavaa koulutusta järjestäville oppilaitoksille ovat maksuttomia (4).

"Tietovastuullinen taho & rekisteri-/tietotuotteet"	Sisältö	"Verkkosaatavuus/julkisuus & koneellinen luettavuus"	Ilmaisuus ja muu mahdollinen maksullinen data
	Säätiörekisteri: Säätiöiden rekisteritietoja, sääntöjä ja tilinpäätöksiä. Rekisteriotteessa on mm. seuraavia tietoja säätiöstä: säätiön nimi, Y-tunnus, kotipaikka ja postiosoite, säätiön tarkoitus, hallitus ja nimen kirjoittaminen. (7) Yhdistysrekisteri: Yhdistysten ja kaupakamarien sekä uskonnollisten yhdistyksien perustietoja ja osoite- sekä nimenkirjoittajätietoja. (8)		Yhdistysrekisteri: Palvelu on maksullinen perustietoja lukuunottamatta. Osoitepoimintoja voidaan myös suorittaa maksua vastaan (9). Vakiomuotoiset tietoluovutukset kunnan ja valtion viranomaisille ovat maksuttomia (4). Muutoin suoritteista (esim. otteet) perittävät maksut ovat omakustannusarvoisesti hinnoiteltuja (yhdistyksiä ja uskonnollisia yhdistyksiä koskevat asiat alle omakustannusarvon), mutta erillaiset rekisteri- sekä tietokantapalvelut ovat liiketaloudellisesti hinnoiteltuja. (6)
PRH & Verohallinto Yritys- ja yhteisötieto-järjestelmä (YTJ)	Tietoja erilaisista yrityksistä ja yhteisöistä. Kaupparekisteriin merkityn yrityksen perustietojen lisäksi palvelussa on tietoa mm. muista rekistereistä, joihin yritys on merkitty (esim. ennakkoperintärekisteri, työnantajarekisteri). Haettavissa myös kaupparekisterin kuulustelutiedot, jotka sisältävät tietoa rekisteröidyistä uusista yrityksistä. Kuulustelutietona julkaistaan myös yritysten tietojen muuttaminen. (10)	Saatavilla julkisesti verkosta (10). <i>Tiedostomuoto:</i> Json. Ladattavissa bulkimuotoisesti verkosta (10).	Maksuton. (10).
Verohallinto Avoim data	Tietoja yhteisöjen ja yhteisötuoksien tuloverotuksen julkisista tiedoista. (11)	Saatavilla julkisesti verkosta (11). <i>Tiedostomuotoja:</i> PDF sekä CSV (11)	Maksuton. (11). Salassapidettävien tietojen vakiomuotoinen luovutus maksutonta viranomaisille (13). Salassa pidettävien tietojen vakiomuotoisen sähköisen luovuttamisen perustaminen ja asiakirjojen luovuttaminen sähköisessä muodossa muutoin kuin vakiomuotoisena on omakustannusarvoisesti hinnoiteltu (12). Toimeksiantosta suoritettavat tutkimus- ja tilastopalvelut ovat liiketaloudellisesti hinnoiteltuja (12).
TRAFI Avoim data	Tietoja liikennekäytössä tai rekisterissä olevista ajoneuvoista, vesikulkuneuvoista ja aluksista. Sisältää esim. tiedot rekisteröinneistä ja teknisistä tiedoista. (1)	Saatavilla julkisesti verkosta. (1) <i>Tiedostomuoto:</i> CSV (1)	Maksuton. (1)
TRAFI Ajoneuvoliikennerekisteri	Tietoja ajoneuvoista ja niiden verotuksesta ja kiinnityksestä, kuljettajista, maakuljetusten turvallisuusneuvonantajista, tieliikenteen valvontalaitteissa käytettävistä korteista, lupaa edellyttävää kuljetustoimintaa varten järjestettävistä kokeista ja myönnettävistä todistuksista. (3)	Ei julkisesti saatavilla. Vaatii yhteydenottoa Trafiin tai sopimusyhtiöihin. Luovutettavan datan laatu riippuu tiedon hyödyntäjän tarkoituksesta ja/tai lakisääteisestä asemasta. (4)	Viranomaisille ja lakisääteistä tehtävää hoitaville yrityksille datan luovutus on maksutonta (5). Muille yrityksille ja yksityishenkilöille luovutus on toteutettu ja hinnoiteltu sopimuskuppaneiden kautta (4) (8). Yksityiselle henkilölle tietoja luovutetaan ajoneuvon rekisteritunnuksen tai valmistenumeron perusteella sähköisessä asiointissa sekä tekstiviesti- ja puhelinalpalveluina (4). Yritykset voivat saada sopimuskuppanien kautta suoramarkkinointitarkoituksiin osoitetietoja ja suorittaa rekisterikyselyjä. (6) Tutkimuskäyttöön luovutettava aineisto vaatii käyttöoikeuden. (7) Tilaukseen tai toimeksiantoon perustuvat tietopalvelusuoritteet ovat liiketaloudellisesti hinnoiteltu ja (2) Kuitenkin ajoneuvojen rekisteröintiin liittyvät suoritteet, ajo-oikeuksiin ja tutkintoihin liittyvät suoritteet sekä lupiin ja valvontaan liittyvät suoritteet, ovat omakustannusperusteisesti hinnoiteltuja (2).
Tilastokeskus Avoim data	Maksuttomat tietokannat sisältävät runsaasti keskeistä perustietoa Suomesta ja maailmalta. Tilastokeskus rakentaa valmiita rajapintoja avoimen datan käyttäjille. Rajapintojen luominen on aloitettu StatFin-tietokannasta. Tilastokeskuksen avoin data -sivuilla esitellään saatavilla olevia avoimia aineistoja ja niiden rajapintoja. (1)	Saatavilla avoimesti verkosta (API-rajapinta) (3). <i>Tiedostomuotoja:</i> PX, CSV, XLSX ja myös muita. Poiminnat mahdollisia. (2).	Maksuton. (2)
Tilastokeskus Maksullinen data	Maksullisissa tietokannoissa on pidempiä aikasarjoja ja yksityiskohtaisempia tietoja eri aihealueista (3). Tilastokeskuksen tutkijapalvelut tarjoaa yksikötason aineistoja (i. mikroaineistoja) tieteellisiin tutkimuksiin ja tilastollisiin selvityksiin. (4)	Lisenssejä ostettavissa sopimuksella, jolla saa ohjelman sekä tunnuksen, joilla voi hyödyntää tilastoja (3). Mikroaineistot vaativat käyttöluvan. Toimintavaihtoehtoina etäkäyttöpalvelu, Tilastokeskuksen tutkimuslaboratorio tai otonaineistojen luovutus tutkijalle (4). <i>Tiedostomuotoja:</i> PX, CSV, XLSX ja muita. Mikroaineistojen tiedostomuodot: excel, csv, sas, txt, stata ja spss	Maksullisiin tietokantoihin on ostettavissa käyttöoikeuslisenssejä, joiden hinnoittelu vaihtelee käytön laajuuden mukaisesti (5). Mikroaineistojen luovuttamiseen tarvitaan käyttö lupa. Toimeksiantannon hinta määräytyy sen laajuuden, vaativuuden ja aineiston käyttömuodon perusteella (4). Tilastokeskuksen kaikki maksulliset suoritteet hinnoitellaan liiketaloudellisin perustein. (7)

"Tietovastuullinen taho & rekisteri-/tietotuotteet"	Sisältö	"VerkkoSaatavuus/julkisuus & koneellinen luottavuus"	Ilmaisuus ja muu mahdollinen maksullinen data
Ilmatieteen laitos Ilmatieteen laitoksen avoin data	Ilmatieteen laitoksen tietoaineistot voidaan jakaa reaaliaikaisiin havaintoihin, aikasarjoihin sekä ennusteisiin. Aineistoissa erilaisia sää-, meri- ja ilmastohavaintoja, joita avataan vaihteittain lisää sitä mukaa, kun aineistot ovat teknisesti valmiita avattaviksi. (1) (2)	Julkisesti saatavilla verkosta, mutta vaatii rekisteröitymisen. Aineistot avattu kone-luettavassa, digitaalisessa muodossa ja aineistoja avataan vaihteittain lisää sitä mukaa, kun aineistot ovat teknisesti valmiita avattaviksi. (2)	Maksuton. Tietoaineistojen räätälöinnistä, palveluiden tuottamisesta sekä muiden kuin avattavien tietoaineistojen käytöstä peritään maksu (4). Sää-, ilmasto- ja meripalvelusuuritteet, jotka tehdään tilauksesta tai pidetään tietopalveluna asiakkaiden saatavina ovat liiketaloudellisesti hinnoiteltuja, erityis-palvelut tietuille viranomaisille on hinnoiteltu omakustannusarvon mukaan (3).
Luke Luonnonvarakeskuksen avoin data	Luken tilastointialueet ovat: 1) maatalouden rakenne, tuotantomenetelmät ja -panokset, peltokasvi-, puutarha- ja kotieläintuotanto, tuotannon vaikutukset ympäristöön sekä maataloustuotteiden hinnat. 2) Metsien taloudellinen hyödyntäminen, puumarkkinoiden toiminta, metsien hoito ja metsien suojeleminen. 3) Kalastus, vesiviljely, kalatalous ja kalastus-tuotteiden markkinointi. 4) Elintarviketurvallisuus ja ravintotase. (1)	Julkisesti saatavilla verkosta (3). Tunnistetietoja sisältävän datan tiedonluovutus on luvanvaraista (2). <i>Tiedostomuotoja:</i> Mm. xml, xls, json (4)	Maksuton. Tunnistetietoja sisältävä data on luvanvaraista sekä maksullista ja voidaan luovuttaa tutkimustarkoitukseen tai toisille viranomaisille (2). Tilastutkimukset sekä tutkimus- ja tilastoaineistot, tietopalvelut, tutkimus- ja muiden aineistojen muokkaaminen sekä siitä syntyvät suoritteet ja tietojärjestelmien, ohjelmien ja vastaavien suoritteiden sekä tietojärjestelmissä olevien tietoaineistojen käyttöoikeuksien ja tietoaineistojen tekninen luovuttaminen tilauksesta ulkopuolisille ovat liiketaloudellisin perustein hinnoiteltuja (5).
Liikennevirasto Liikenneviraston avoin data	Liikennevirasto tuottaa ja julkaisee liikenteeseen ja väyliin liittyviä karttoja, tilastoja, julkaisuja, tietoaineistoja ja avointa dataa. Aineisto on tarkoitettu kaikkien kansalaisten ja yhteiskunnan eri toimijoiden käyttöön. Viraston avoin data sisältää liikenteen reaaliaikaistati-lannetietoa ja staattista tietoa liikenne-verkoista. Myös muita avoimia aineistoja joukkoliikenteen aikatauluista ja reiteistä sekä henkilöliikennetutkimuksesta ja liikenteen ennustamisesta. (1)	Julkisesti saatavilla verkosta (2). <i>Tiedostomuoto:</i> SHAPE (ainakin). Liikennevirasto julkaisee myös julkisen liikenteen suoritetilaston, henkilöliikennetutkimukseen liittyviä tilastoja sekä tilastoja liikenne- ja hirvieläinonnettomuuksista maanteillä, tasoristeysonnettomuuksista sekä turvallisuuspoikkeamista pdf-muodossa (3).	Maksuton. (2) Muita aineistoja saatavilla ainakin tutkimuskäyttöön lupaprosessin kautta (4). Toimeksiantoon perustuvat tilastopalvelut sekä muu tietopalveluiden- ja aineistojen myynti on hinnoiteltu liiketaloudellisin perustein (5). Mutta kuitenkin valtion maksuperustellaista tarkoitettuja muita suoritteita, joissa laitosella on tosiasiallinen yksinoikeus suoritteen tuottamiseen ja joissa maksu voidaan määrätä siten, että se vastaa tuotteen omakustannusarvoa, ovat tilauksesta laadittavat tilastot, koodistot ja tietotuotteet sekä laajat tietopalvelutehtävät (5).
THL Terveyden ja hyvinvoinnin laitoksen avoin data	Avoimena datana ovat saatavissa tilasto- ja indikaattoripankki Sotkanetin, elintarvikkeiden koostumustietokanta Finelin, sosiaali- ja terveydenhuollon toimipaikkarekisteri TOPI:n sekä Tartuntatautirekisterin tiedot. Lisäksi avointa dataa hoitoonpääsystä erikoissairaanhoidossa, lastensuojelun käsittelyajoista, hedelmöityshoidoista, raskaudenkeskeytyksistä, steriloinneista, synnytyksistä sekä toimeentulotuesta. (1)	Julkisesti saatavilla verkosta (1). Kuitenkin esim. valtakunnallisten sosiaalihuollon sekä terveydenhuollon rekisterien tiedot ovat salassa pidettäviä. THL voi kuitenkin yksittäistapauksessa antaa luvan kyseisten rekisterien ja asiakirjojen tieteelliseen tutkimuskäyttöön. (3) <i>Tiedostomuoto:</i> json, csv (2)	Maksuton. Salassapidettävien tietojen osalta tutkimusaineistopöytäkirjasta perittävä maksu määräytyy siihen käytetyn työajan perusteella, kun käyttölupa on myönnetty (4). Tämä on hinnoiteltu omakustannearvoisesti ja koskee tilauksesta laadittavia tilastoja, koodistoja, tietotuotteita ja laajoja tietopalvelutehtäviä. Lisäksi salassa pidettävien aineistojen käyttöluopapäätöksiä maksuissa noudatetaan omakustannepöytäkirjasta hinnoiteltua. (5)
KELA Avoimen datan tietokanta Kelasto	Kelastoon on koottu keskeinen tilastotieto Kelan hoitamasta sosiaaliturvasta: etuuksien saajista, summista sekä etuuksien hakemuksista ja ratkaisuisista. (1)	Julkisesti saatavilla verkosta. Lisäksi tietoluovutuksia suoritetaan laajemmista aineistoista valikoidusti (1). <i>Tiedostomuoto:</i> xls (1)	Maksuton. Avoimen datan lisäksi luovutettavista tiedoista Kela ei laskuta itse tiedoista, vaan muista tietoluovutukseen liittyvistä kustannuksista. Kelalle tulee etuustietojen kyselypalveluiden lisäksi erilaisia ulkopuolisia tieteellisen tutkimuksen aineistopyyntötietoja. Näistä tietoluovutuksista laskutetaan ainoastaan IT-työstä. (2)

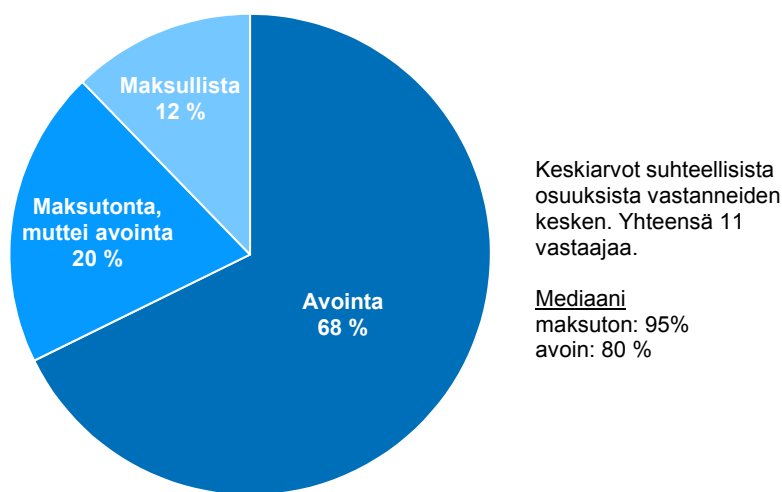
Taulukossa suluissa olevat numerot viittaavat kunkin organisaation kohdalla käytettyihin lähteisiin, jotka löytyvät liitteestä 4.

2.2 Perustietovarantojen avaaminen ja käyttö

Keskeisten rekisteritietojen hallinnoivien organisaatioiden tietojen avoimuutta ja maksuttomuutta koskeva kysely toteutettiin sähköpostitse elo-syyskuussa 2016. Verkossa täytettävä kyselylomake lähetettiin seuraaville organisaatioille: Väestörekisterikeskus (VRK), Patentti- ja rekisterihallitus (PRH), Verohallinto, Maanmittauslaitos (MML), Trafi, Tilastokeskus, Ilmatieteenlaitos, Luonnonvarakeskus, Suomen Ympäristökeskus (SYKE), Liikennevirasto, Kansaneläkelaitos ja Terveyden ja hyvinvoinnin laitos (THL). Kyselylomake oli pääosin sama kuin Etlan joulukuussa 2009 toteuttamassa kyselyssä käytetty julkisten tietovarantojen maksullisuutta kartoittava kysymyslomake. Muutamia täydentäviä kysymyksiä lisättiin. Kyselylomake on liitteessä 2. On huomattava, että tässä luvussa esitetyt tiedot kuvastavat tilannetta syksyllä 2016 ja voivat olla raportin julkaisuajankohtana jo osin vanhentuneita.

Kaikkiin kysymyksiin saatiin vastaukset kahdeksalta organisaatiolta. Organisaatioilta kysyttiin ensinnäkin niiden tarjoamien tietotuotteiden maksuttomuudesta ja avoimuudesta niiden tietojen osalta, joiden avaamiselle ei ole lakisääteistä estettä. Vastaukset osoittavat, että datan avaaminen on varsin pitkällä keskeisissä rekisteritietoja hallinnoivissa organisaatioissa. Keskimäärin liki 90 prosenttia (mediaanilla mitattuna 95 prosenttia) kysymykseen vastanneen 11 rekisterinpitäjän tietotuotteista oli maksuttomia. Tiedoista, joiden avaaminen on lakisääteisesti mahdollista, avointa oli keskimäärin 69 prosenttia (mediaanilla mitattuna 80 prosenttia).

Kuvio 2.1 Maksuttomien ja avoimeksi dataksi laskettavien (avoimesti lisensoitu, koneluettava ja maksuton) tietotuotteiden osuus kaikista tietotuotteista, joiden avaamiselle ei ole lakisääteistä estettä

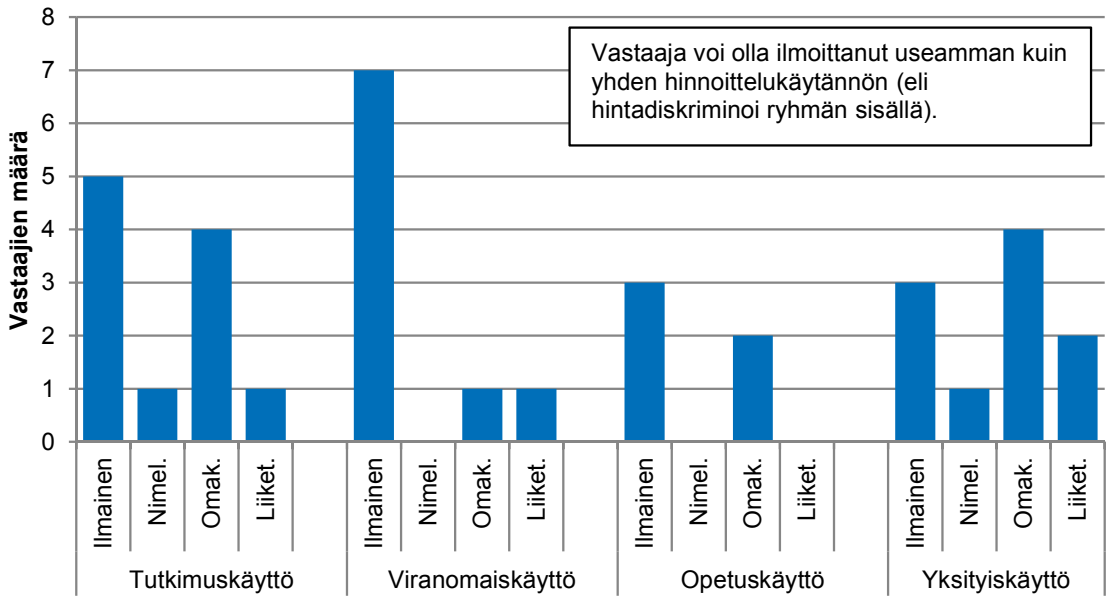


Vastaajia pyydettiin myös arvioimaan sitä, kuinka suuri osuus organisaation tietotuotteista on hinnoiteltu valtion maksuperustelain 150/92 hinnoittelukäytäntöjen mukaisesti. Maksuperustelaisissa julkisen sektorin suoritteet jaetaan maksuttomiin (väljästi tulkiten yleishyödylliset suoritteet), omakustannehintaisiin ja liiketaloudellisesti hinnoiteltaviin. Omakustannehintaisia ovat sellaiset valtion viranomaisen tarjoamat suoritteet, joiden kysyntä perustuu lakiin tai asetukseen ja jonka tuottamiseen viranomaisella on tosiasiallinen yksinoikeus. Omakustannehinta tarkoittaa sitä, että tietotuotteen tuottamisen kaikki kustannukset katetaan eli käytännössä tietotuotteen hinta määräytyy sen käyttäjämäärän perusteella keskimääräisen kustannuksen mukaan. Maksuperustelain puitteissa julkisen sektorin organisaatio pystyy useimmiten itse määrittämään tietotuotteidensa hinnoittelutyyppin.

Vain yksi vastanneista organisaatioista ilmoitti käyttävänsä nimellistä irtiottokuluihin perustuvaa hinnoittelua. Viisi vastaajaa kahdeksasta ilmoitti organisaationsa hinnoittelevan tietotuotteitaan omakustannehintaan perustuen ja myös viisi käyttävänsä liiketaloudellista hinnoittelua. Hintadiskriminoinnin käyttö eli tietotuotteiden tarjoaminen eri käyttäjärhyille erilaisin hinnoitteluperustein oli yleistä.

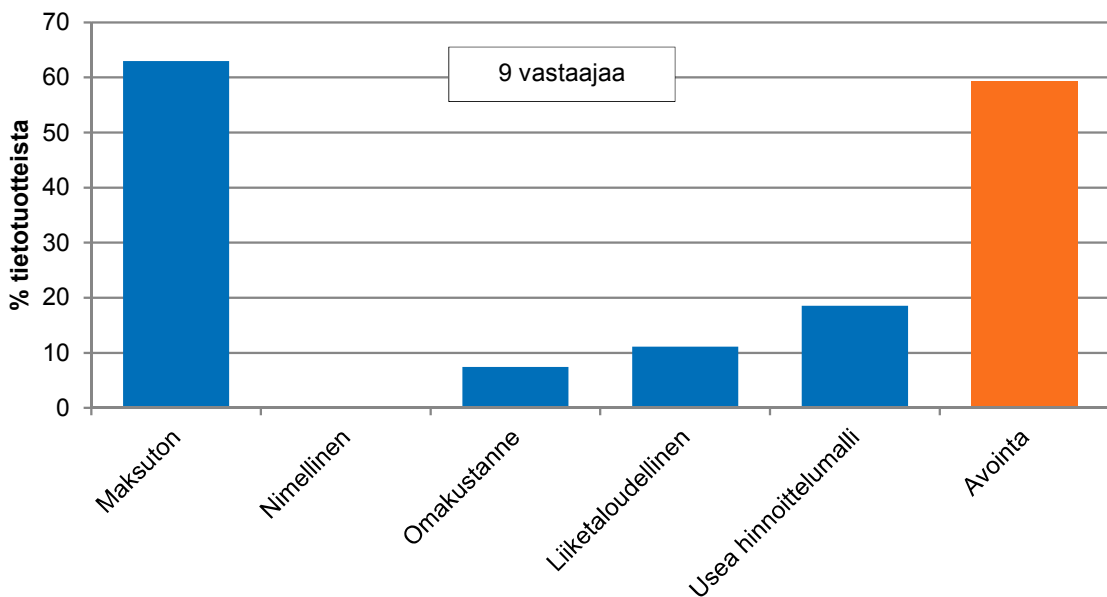
Erityisesti viranomaiskäyttöön tietoja luovutetaan maksutta. Muuten hinnoittelukäytännöt jakautuvat melko tasaisesti. Kustannussyyt ovat edelleen merkittävin este sille, ettei vielä avaamattomia dataa voida tarjota maksuttomina; dataa voitaisiin tarjota nykyistä enemmän maksuttomana, mikäli kustannukset katettaisiin budjettirahoitteisesti.

Kuvio 2.2 Maksullisten tietotuotteiden hintadiskriminointi



Vastaajia pyydettiin arvioimaan kolme käytetyintä tietotuotetta tai tietotuoteryhmää ja kommentoimaan näiden hinnoittelua. Maksuttomia käytetyimmistä tietotuotteista oli 63 prosenttia ja niistä avointa dataa noin 59 prosenttia. Mikäli vastaaja ilmoitti tietotuotteiden hinnoittelussa käytettävän hintadiskriminointia, se on tässä niputettu ”usea hinnoittelumalli” -kohtaan. Näissä oli ilmoitettu erilaisina yhdistelmiä maksutonta, omakustanneperusteista ja/tai liiketaloudellista hinnoittelua. Irtiottokustannuksiin perustuvaa hinnoittelumallia hyödyntäviä tietotuotteita ei käytetyimpien tietotuotteiden joukosta löytynyt.

Kuvio 2.3 Suosituimpien tietotuotteiden hinnoittelu

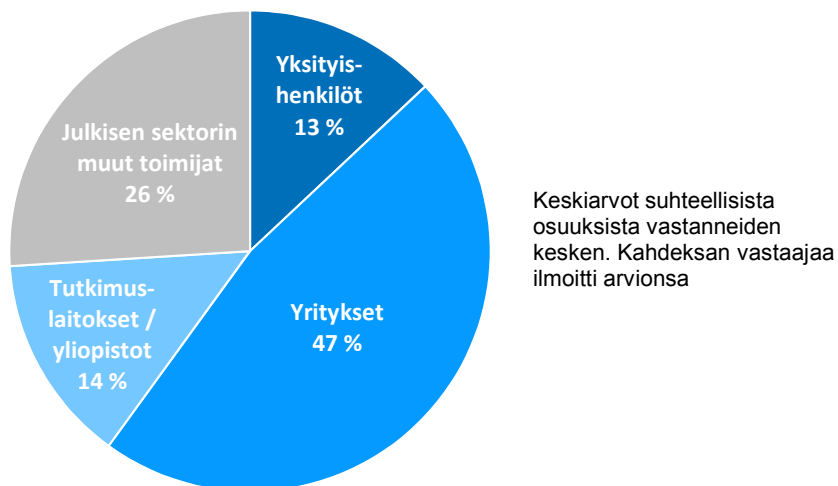


Yhdeksän rekisterinpitäjää vastasi kysymykseen koskien niiden tietotuotteista vuonna 2015 saamia tuloja. Niiden yhteenlasketut tietotuotetulot vuodelta 2015 olivat hieman yli 34 miljoonaa euroa. Kahdella rekisteripitäjällä ei ollut lainkaan tietotuotetuloja, kun taas kolmen organisaation osalta tietotuotetuloja kertyi 9–10 miljoonaa euroa. Organisaation tietotuotteiden tuottamisen kustannukset raportoi seitsemän organisaatiota. Ne olivat yhteensä liki 29 miljoonaa euroa. Vertailukelpoista aineistoa vuoden 2009 joulukuussa kerätyistä tiedoista saatiin tietotuotetulojen ja -kustannusten osalta kuudesta organisaatiosta. Näiden rekisterinpitäjäorganisaatioiden yhteenlasketuissa tietotuotetuloissa oli tapahtunut jonkin verran nousua, mutta tietotuotekustannuksissa puolestaan laskua. Tietotuotteista kertyneiden tulojen ja niiden tuottamisen aiheuttamien kustannusten ero oli verrattain pieni vuonna 2015, kun taas vuonna 2008 kustannukset ylittivät selkeästi tulot. Vuonna 2008 tietotuotteiden tuotannon kustannukset olivat kuudessa molempiin kyselyihin vastanneissa organisaatioissa yli kaksikertaiset verrattuna tietotuotteista saatuihin tuloihin. Toisin sanoen, tietotuotteiden tuottamisesta aiheutuvat kustannukset ovat laskeneet merkittävästi ainakin osassa rekisterinpitäjäorganisaatioita.

Vastaukset viittasivat kuitenkin siihen, että tietotuotteista saatujen tulojen ja kustannusten arviointi on monen organisaation osalta hankalaa. Yksiselitteistä tapaa laskea tietotuotetuloja- ja kustannuksia ei ole. Tämän takia ainakin joissakin tapauksissa vastaajan subjektiivinen tietotuotetulojen ja -kustannusten arviointitapa (eli se mitä luetaan tietotuotteiden tuloihin ja kustannuksiin) vuoden 2016 kyselyssä voi poiketa organisaatiossa vuoden 2009 kyselyyn vastanneen arviointitavasta. Tämä heikentää vuoden 2015 tietotuotetulojen ja -kustannusten vertailtavuutta sekä organisaatioiden välillä että yksittäisten organisaatioiden vuosien 2008 ja 2015 tietojen välillä. Yllä esitettyjä lukuja voidaankin pitää vain suuntaa antavina.

Käytetyimpiin tietotuotteisiin (taulukko 2.2) sisältyy erityisesti ajantasaisia ja historiallisia tilastotietokantoja, kuten erilaisia rekisteritietoja kansalaisista ja yrityksistä sekä paikkatietoaineistoja. Myös reaaliaikaisen paikkatiedon siirtämisen ja hyödyntämisen tai asiakasrekisterien tietojen päivittämisen mahdollistavat rajapinnat ovat suosittuja, vaikka nämä tiedot ovatkin osaksi luvanvaraisia. Suosituimpien tuotteiden osalta yksityishenkilöitä ja näiden omaisuutta koskeva tieto on pääsääntöisesti maksullista (omakustanneperusteisesti tai liiketaloudellisesti hinnoiteltua) ja voi rajoittua loppukäyttäjän aseman perusteella yksittäisiin otteisiin tai luvanvaraisiin poimintoihin.

Kuvio 2.4 Tietotuotteiden käyttäjäryhmien suhteelliset osuudet



Taulukko 2.2 Suosituimmat tietotuotteet

	Tietotuotteen 1 kuvaus	Tietotuotteen 2 kuvaus	Tietotuotteen 3 kuvaus
Suomen ympäristökeskus	Paikkatietoaineistot, (pohjavesialueet, luonnonsuojelualueet ja maanpeiteaineistot) Avoin	Ympäristötietojärjestelmien avoimet rajapinnat (mm. vedenlaatutiedot ja kasviplanktonitiedot) Avoin	Ympäristötietojärjestelmät (hydrologia, vesienhoidon suunnittelu/toteutus, pohjavesien suojele) [Avoin, vaatii rekisteröitymisen]
Verohallinto	Verotuksen tilastot	Tuloverotuksen julkiset tiedot	Tuloverotuksen julkiset tiedot (yhteisöt) Avoin
Tilastokeskus	StatFin (maksuton tilastotietokanta, joka sisältää Suomea koskevaa keskeistä tilastotietoa) Avoin	Paavo – Postinumeroalueittainen avoin tieto (tilasto- ja kartta-aineistoja.) Avoin	Rajapintapalvelut. Paikkatietoaineistojen katselua ja siirtämistä tarvittaessa omaan sovellukseen. Avoin
Liikenteen turvallisuusvirasto Trafi	Ajoneuvoliikennerekisterin tietojen luovuttaminen suoramarkkinoitiin	Ajoneuvoliikennerekisterin tietojen luovutus yrityksille ja kansalaisille suunnattuihin kyselypalveluihin	Ajoneuvotietojärjestelmän suorakäyttö viranomais- ja yritysten lakisääteisiin tehtäviin
Liikennevirasto	Paikkatietojen rajapinta- ja tiedostolautauspalvelu. (aineistoja tie-, rata- ja meripaikkatietojen osalta. Digiroad-aineisto. Merikartta-aineistot rajapintapalveluna) Avoin	Digitraffic-reaaliaikatieojen rajapintapalvelu (tie-, rata- ja meripuolen reaaliaikaisia liikennetietoja ja näiden historiatietoja) Avoin	Navigointiin virallisia elektronisia ja painettuja merikarttoja koskeva merikarttojen kustannus- ja jakelutoiminta sekä merikartta-aineiston sopimusperustainen lisensointi
Väestörekisterikeskus	Asiakasrekisterin päivityspalvelu (viranomaisien ja yritysten henkilöasiakkaiden tietojen jatkuvan ylläpidon VTJ:n ajantasaisilla tiedoilla. Edellyttää käyttöoikeusluvan)	Aineistopalvelu (satunnaisotannat ja poiminnat väestötietojärjestelmästä. Edellyttää käyttöoikeusluvan. Rakennusten sijainti ja osoitteet avointa dataa)	Väestötietojärjestelmän rajapintapalvelu asiakasorganisaatioille (edellyttää käyttöoikeusluvan)
Luonnonvarakeskus	Valtakunnan metsien inventoinnin karttatasot Avoin	Luonnonvaratalouden tilastot Avoin	Riistahavainnot.fi -palvelu (tietoa hirvien ja suurpetojen kannoista sekä havainnoista)
Ilmatieteen laitos	Säähavainnot (lämpötila, kosteus, tuuli, paine, sade jne.) Avoin	Sääennustedata (sääennustemallin piste-ennusteita ja hilakenttiä). Avoin	Sääatutkahavainnot (sateen voimakkuus, tutkaohjastuvuustekijä, tuulen säteisyys sekä sadekertymät 1,12 ja 24 h) Avoin
Maanmittauslaitos	Kiinteistötietojärjestelmän tuotteet	Maastotietojärjestelmän tuotteet Avoin	Kauppahintarekisterin tuotteet

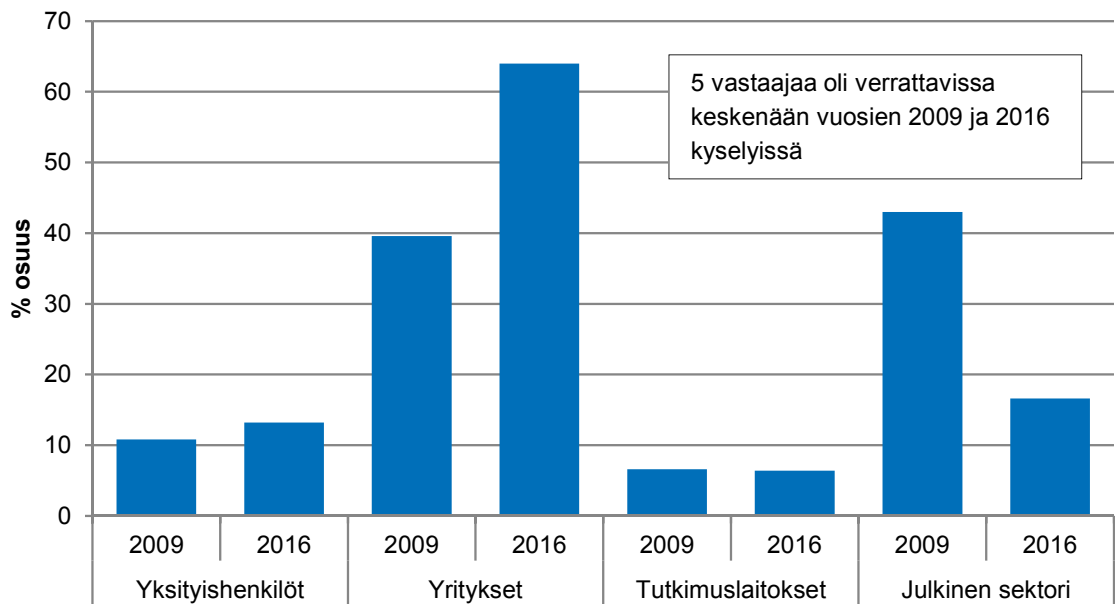
Vastaajien arvioiden mukaan tietotuotteiden suurin käyttäjäryhmä ovat yritykset: yritysten osuuden arvioitiin olevan keskimäärin liki puolet (perustuen käytettyjen tietotuotteiden lukumäärään). Julkisen sektorin muut organisaatiot olivat toiseksi merkittävin julkisen tiedon käyttäjäryhmä, yksityishenkilöiden osuus oli noin 13 prosenttia ja tutkimuslaitosten ja yliopistojen 14 prosenttia käytetyistä tietotuotteista.

Etlan joulukuussa 2009 toteuttaman julkisen tiedon hinnoittelua ja tarjontaa kartoittavan kyselyn aineistoa³ hyödyntäen voitiin vertailla käyttäjäryhmien osuuksien muutosta viiden vastaajan osalta. Näiden vastaajien mukaan julkisen sektorin suhteellinen osuus on pienentynyt eniten, kun yksityishenkilöiden ja erityisesti yritysten osuus on kasvanut. Tutkimuslaitosten ja yliopistojen osuus ei ole muuttunut merkittävästi vuoden 2008 ja 2015 välillä.

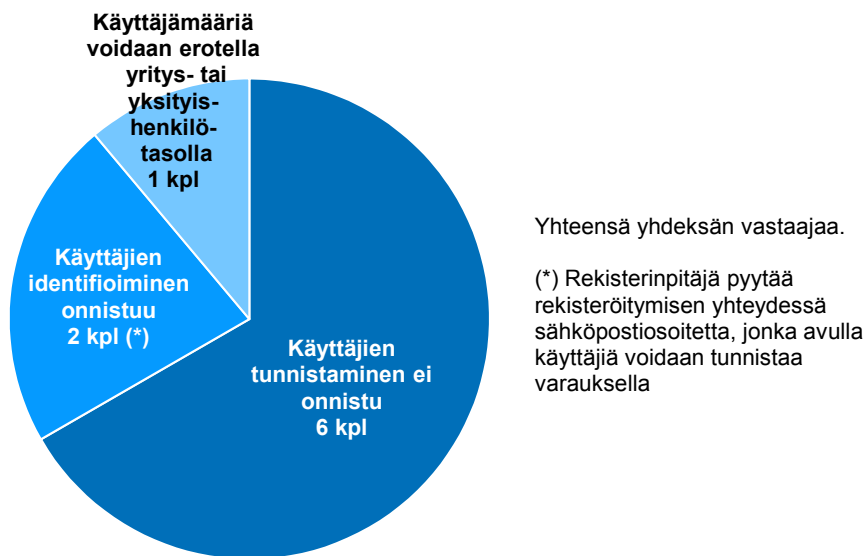
Avoimen datan vaikuttavuuden tutkimuksen näkökulmasta datan käyttäjien anonymisuus hankaloittaa työtä. Suurin osa vastaajista raportoi, ettei organisaation avoimen datan käyttäjiä pystytä tällä hetkellä tunnistamaan ja lopuillakin tunnistaminen perustui lähinnä rekisteröinnin yhteydessä vaadittavaan sähköpostiosoitteeseen. Usein mainituksi syyksi nousi avoimen datan periaate, ettei käyttäjiä tunnisteta eikä datan käyttö vaadi rekisteröitymistä.

³ Kyselyn tulokset on raportoitu Kosken (2010) julkaisussa "Julkisen tiedon hinnoittelu ja tarjonta Suomessa. Valtioneuvoston kanslia. Kasvutyöryhmän taustamuistio."

Kuvio 2.5 Tietotuotteiden käyttäjäryhmien suhteelliset osuudet 2009 ja 2016 kyselyissä



Kuvio 2.6 Avoimen tiedon käyttäjien tunnistettavuus



3. AVOIMEN DATAN TALOUDELLINEN VAIKUTTAVUUS

Heli Koski – Mika Pajarinen

Avoimen datan taloudellisen vaikuttavuuden tutkimisen suurimpana haasteena on tällä hetkellä systemaattisen tiedonkeruun puute avoimen datan hyödyntämisestä. On kuitenkin huomattava, että avoin data muodostaa vain pienen osan eri käyttäjäryhmien hyödyntämästä datasta. Aineistoanalyysimme osoittaa, että sekä avointa dataa että massadataa innovaatio-toiminnassaan hyödyntävät yritykset tekevät muita yrityksiä huomattavasti todennäköisemmin uusia tuoteinnovaatioita. Dataa uusien palveluiden ja/tai tuotteiden kehittämisessä hyödyntäneet yritykset ovat kasvaneet muita yrityksiä nopeammin joillakin tietointensiivisillä sektoreilla. Datatalouden mekanismien ymmärtämiseksi tarvittaisiin tutkimustietoa yritysten datan käytöstä ja taloudellisista vaikutuksista laajemmin. Aihealueen aineistoanalyysiin perustuvan tutkimuksen mahdollistamiseksi tulisi kehittää datan käyttöä kuvastavaa tilastoaineistojen keräämistä.

3.1 Minkälaisia mittareita tarvitsemme, onko meillä dataa?

Avoimen datan taloudellista vaikuttavuutta voidaan mitata useilla eri tasoilla kansantalouden tasolta aina kansalaistasolle asti. Tässä käsitelty avoimen datan vaikuttavuusanalyysissa tarvittava mittaristo pohjautuu Kosken (2015) esitutkimusraporttiin. Mittaristo koostuu kahdesta osa-alueesta: odotettuja vaikutuksia (hyötyjä) kuvaavista mittareista ja avoimen datan käyttöä kuvaavista mittareista. Avoimen datan vaikuttavuuden mittarit on jaettu datan hyödyntäjäryhmien mukaan seuraavasti: i) yritykset, ii) kansalaiset, iii) julkinen sektori ja iv) aggregaattina koko kansantalouden taso.

Avoimen tiedon odotettuja vaikutuksia ovat muun muassa yritysten osalta niiden avoimeen dataan perustuvat uudet tuotteet ja palvelut, toiminnan tehostaminen uusilla, dataperusteisilla päätöksentekojärjestelmillä sekä uuden yritystoiminnan syntyminen. Parhaimmillaan dataperusteisista innovaatioista syntyy kasvua ja tuottavuuden paranemista. Kansalaisille merkittävimmät avoimen datan taloudelliset hyödyt syntyvät aiemmin maksullisen tiedon muuttumisesta maksuttomaksi sekä avoimen datan ja sen avulla luotujen sovellusten mahdollistamasta ajankäytön tehostumisesta. Julkisen sektorin organisaatiot voivat myös tehostaa palveluitaan ja saavuttaa kustannussäästöjä avoimen datan avulla. Kosken (2015) raportissa keskustellaan tarkemmin avoimen tiedon odotetuista vaikutuksista eri käyttäjäryhmille.

Taloudellisen vaikuttavuuden analyysi vaatii paitsi menetelmällistä osaamista (kts. luku 3.2), myös huolellisesti rakennettuja vaikuttavuusmittareita sekä pidemmältä ajanjaksolta kerättyä tietoa datan käytöstä ja mitattavista vaikutuksista. Tässä luvussa on arvioitu olemassa olevia potentiaalisia tietolähteitä ja tilastoaineistojen kattavuutta Kosken (2015) mittariston pohjalta. Valmiita, hyödynnettäviä odotettujen vaikutusten mittareita olisi saatavilla kansantalouden tasolla sekä yritystasolla kasvun ja tuottavuuden mittarien suhteen, tosin yritystason mittarit ovat maksullisia. Muiden mittarien osalta voitaisiin käyttää erilaisia lakisäätteisiä kyselytutkimuksia, joiden toteutusväli on yhdestä kahteen vuotta.

Usein julkisen sektorin avointa dataa käytetään yhdessä muiden datojen (esim. yrityksen ostamien tai omien datojen) kanssa, mikä muodostaa omanlaisensa harmaan alueen avoimen

Taulukko 3.1 Odotettujen vaikutusten mittarit, saatavuus ja tilastojen kattavuus

Mittattava taso	Odotettu vaikutus	Vaikutuksen mittarit	Potentiaalisia tietolähteitä	Tilaston kattavuus
Yritykset	Uudet tuotteet ja palvelut	Avoimeen dataan perustuvien tai sitä osittain hyödyntävien uusien tuotteiden ja palveluiden määrä. Avoin data palveluiden/tuotteiden tuottamien tulojen osuus liikevaihdosta	Kyselytutkimukset; kysymysten lisääminen esim. <i>Tietotekniikan käyttö yrityksissä</i> -kyselyyn ja CIS -kyselyyn (kysymykset voisivat koskea mm. mihin yritykset käyttävät dataa ja miten ne hyötyvät siitä tai kuinka tärkeää liiketoiminnan kannalta avoin data on)	<i>Tietotekniikan käyttö yrityksissä</i> -kyselyyn vuonna 2015 vastasi hyväksytyksi 3 070 yritystä. CIS-kyselyssä tiedot kerätään kaikilta yli 250 henkilöä työllistäviltä yrityksiltä teollisuudessa, kaivostoiminnassa ja louhinnassa, sähkö-, kaasu- ja lämpöhuollossa ja vesi- ja jätehuollossa sekä valituilla palvelualoilla. Tätä pienemmistä yrityksistä poimitaan otos. <i>Tietotekniikan käyttö yrityksissä</i> -kysely suoritetaan vuosittain, CIS-kysely kahden vuoden välein.
	Uudet yritykset	Uusien avointa dataa hyödyntävien yritysten lukumäärä		
	Kasvu	Liikevaihdon muutos, työntekijöiden määrän muutos	1) Tilastokeskuksen yritysrekisteri 2) Asiakastiedon tilinpäätöstietokanta	1) Kaikki yritykset, yhteisöt ja yksityiset elinkeinonharjoittajat, jotka ovat liiketoiminnasta arvonlisäverovelvollisia tai toimivat työnantajina. Yritysrekisterin tiedot päivittyvät 1–10 kuukauden viiveellä. 2) 150 000 suomalaisen yrityksen tilinpäätöstiedot. Julkaistaan kaksi kertaa vuodessa.
	Tuottavuus	Arvonlisä per työntekijä tai tehdyt työtunnit		
Kansalaiset	Säästetty aika ja raha	Käytetyn ilmaisen tiedon määrä x aiempi hinta Arvio aikasäästöistä / saavutetuista hyödyistä	Kyselytutkimukset; kysymysten lisääminen <i>Väestön tieto- ja viestintätieteiden käyttö</i> -kyselyyn (kysymykset voisivat koskea mm. mihin kansalaiset käyttävät dataa ja miten he hyötyvät siitä)	<i>Väestön tieto- ja viestintätieteiden käyttö</i> -kyselyyn vuonna 2015 vastasi 4 850 henkilöä tai kotitaloutta. Kysely suoritetaan vuosittain.
Julkinen sektori	Kustannussäästöt	Käytetyn työajan muutos x työajan hinta	1) Automatisoitu tietojen ja palvelujen tarjonnan seuranta 2) Kyselytutkimukset; <i>Tietoja valtion tietohallinnosta</i> -kysely ja siihen lisättävät kysymykset	Ministeriöt ja näiden alaiset virastot ja organisaatiot. <i>Tietoja valtion tietohallinnosta</i> -kysely suoritetaan vuosittain.
	Tuottavuus	Tuotetut palvelut suhteessa käytettyihin resursseihin		
Kansantalous	Kasvu	Arvonlisän eli BKT:n muutos	Tilastokeskuksen tilastotietokannat	
	Tuottavuus	BKT asukasta kohden		

datan vaikuttavuuden tutkimisessa. Toisin sanoen voi olla vaikeaa arvioida, missä määrin avoin data on vaikuttanut tiettyyn mittariin, eli missä kohtaa sen vaikutus on kriittistä esimerkiksi uuden tuotteen tai palvelun syntymisessä ja missä kohtaa vain täydentävää.

Taulukko 3.2 Avoimen datan käytön mittarit ja näiden saatavuus

Mittattava taso	Avoimen datan käytön mittarit	Potentiaalisia tietolähteitä	Tilaston kattavuus
Yritykset	Kuinka paljon ja mitä dataa yritykset käyttävät	1) Tiedon keruu avoimen tiedon käyttöportaaleista 2) Kyselytutkimukset; kysymysten lisääminen esim. <i>Tietotekniikan käyttö yrityksissä</i> -kyselyyn ja CIS -kyselyyn	1) Potentiaalisesti kaikki datan tunnistettavissa olevat yritystason käyttäjät 2) <i>Tietotekniikan käyttö yrityksissä</i> -kyselyyn vuonna 2015 vastasi hyväksytyksi 3 070 yritystä. CIS-kyselyssä tiedot kerätään kaikilta yli 250 henkilöä työllistäviltä yrityksiltä teollisuudessa, kaivostoiminnassa ja louhinnassa, sähkö-, kaasu- ja lämpöhuollossa ja vesi- ja jätehuollossa sekä valituilla palvelualoilla. Tätä pienemmistä yrityksistä poimitaan otos. Kysely suoritetaan vuositasolla.
Kansalaiset	Kuinka paljon ja mitä dataa kansalaiset käyttävät Kuinka paljon avoimeen dataan pohjavia sovelluksia kansalaiset käyttävät	1) Tiedon keruu avoimen tiedon käyttöportaaleista 2) Kyselytutkimukset; kysymysten lisääminen <i>Väestön tieto- ja viestintätieteiden käyttö</i> -kyselyyn	1) Potentiaalisesti kaikki datan tunnistettavissa olevat yksityishenkilötason käyttäjät 2) <i>Väestön tieto- ja viestintätieteiden käyttö</i> -kyselyyn vuonna 2015 vastasi 4 850 henkilöä tai kotitaloutta. Kysely suoritetaan vuositasolla.
Julkinen sektori	Kuinka paljon ja mitä dataa hyödynnetään julkisen sektorin organisaatioissa	1) Tiedon keruu avoimen tiedon käyttöportaaleista? 2) Kyselytutkimukset; <i>Tietoja valtion tietohallinnosta</i> -kysely ja siihen lisättävät kysymykset	1) Potentiaalisesti kaikki datan tunnistettavissa olevat julkisen sektorin organisaatiot 2) Ministeriöt ja näiden alaiset virastot ja organisaatiot. Kysely suoritetaan vuositasolla.
Kansantalous	Aggregaattitilastot avoimen tiedon käytöstä, mitä tietoja käytetään eniten	Avoimen tiedon käyttöportaaleista saatavat tiedot	Organisaatio ja/tai tuotekohtaiset avoimen tiedon lataukset

Pääsääntöisesti avoimen tiedon käyttötietojen keräämisessä voidaan harkita kahta lähestymistapaa: i) kyselytutkimuksia tai ii) käyttötietojen suoraa keräämistä avoimen tiedon käyttöportaaleista. Avoimen datan käytöstä ei kerätä tällä hetkellä systemaattista tietoa kumpaakaan tapaa käyttäen. Vaikuttavuusanalyysin näkökulmasta avoimen datan hyödyntämisestä tarvittaisiin mieluiten käyttäjätasolle identifiotavissa olevaa dataa. Tällaista dataa eivät rekisteritietoja hallinnoivat organisaatiot tällä hetkellä juurikaan kerää, mutta edellytykset datan käyttäjien identifiointiin ja rekisteröintiin ovat olemassa. Avoimen datan hyödyntämistä mittaavien kysymysten sisällyttäminen Tilastokeskuksen toteuttamiin lakisääteisiin kyselytutkimuksiin mahdollistaisi avoimen datan käytön pitkäaikaisen tilastoinnin ja seurannan. Lisäksi avoimen datan taloudellista vaikuttavuutta voitaisiin tutkia kvantitatiivisesti kehittyneitä tilastollisia menetelmiä käyttäen, kun kyselyaineistot yhdistettäisiin muihin yritys- tai henkilötason tietokantoihin.

Tilastokeskuksen toteuttamien yritys-kyselyiden osalta on datan käyttöä koskevan tiedon osalta nähtävissä positiivista kehitystä. Datan hyödyntämistä yritysten innovaatiotoiminnassa kartoittava kysymyspatteristo lisättiin Tilastokeskuksen joka toinen vuosi toteuttamaan Yritysten innovaatiotoiminta (Community Innovation Survey, CIS) -kyselyyn vuoden 2015 kyselylomakkeeseen (ks. taulukko 3.3). Vuoden 2015 kysely kartoitti yritysten innovaatiotoimintaa vuosien 2012–2014 ajalta. Yrityksiä pyydettiin arvioimaan datan merkitystä niiden t&k-toiminnassa, uusien tuotteiden ja palveluiden kehittämisessä ja parantamisessa sekä muussa innovaatiotoiminnassa (ts. prosessi-, organisaatio- ja markkinointi-innovaatioiden kehittämisessä). Lisäksi kysyttiin datan käytön merkityksestä yritykselle sen tuotantoprosessin hallinnassa ja markkinoinnissa sekä sitä, kuinka tärkeää big datan (esim. asiakastiedot) myynti ja ostaminen ovat yritykselle. Datan hyödyntämistä kartoittavan kysymyksen sisällyttämistä myös vuoden 2017 syksyllä toteutettavaan Yritysten innovaatiotoiminta 2014–2016 -kyselyyn on esitetty aikasarjatiedon keräämiseksi. Tilastokeskuksen kanssa käydään myös keskusteluja datan hyödyntämistä mittaavien kysymysten sisällyttämiseksi vuosittain toteutettavaan Tietotekniikan käyttö yrityksissä -kyselyyn. Pohjana yritysten datan käyttöä koskevien mittarien keräämiseksi voitaisiin hyödyntää taulukon 3.3 kysymyspatteristoa, muokaten sitä soveltuvin osin. Esimerkkinä kysymyspatteristoon lisättävistä kysymyksistä on keinoälyn rooli datan hyödyntämisessä.

Taulukko 3.3 Miten suuri merkitys yrityksellenne on seuraavilla datan käyttöön liittyvillä seikoilla?

	Merkitys			
	Suuri	Kohtalainen	Vähäinen	Ei koske yritystämme
Datan käyttö tuote-innovaatioiden eli uusien ja olennaisesti parannettujen tavaroiden ja palveluiden kehittämisessä:				
Datan käyttö uusien tavaroiden ja/tai palveluiden kehittämisessä				
a) Big data	r	r	r	r
b) Julkisen sektorin avoin data	r	r	r	r
Datan käyttö tavaroiden ja/tai palveluiden parantamisessa				
a) Big data	r	r	r	r
b) Julkisen sektorin avoin data	r	r	r	r
Datan käyttö muussa innovaatiotoiminnassa				
a) Prosessi-innovaatioiden kehittäminen	r	r	r	r
b) Organisaatio-innovaatioiden kehittäminen	r	r	r	r
c) Markkinointi-innovaatioiden kehittäminen	r	r	r	r
d) Logistiikkaan liittyvien innovaatioiden kehittäminen	r	r	r	r
Datan käyttö tutkimus- ja kehittämistoiminnassa	r	r	r	r
Datan käyttö tuotantoprosessin hallinnassa	r	r	r	r
Datan käyttö markkinoinnissa	r	r	r	r
Big datan (esim. asiakastiedot) myynti muille yrityksille	r	r	r	r
Big datan (esim. asiakastiedot) osto muilta yrityksiltä	r	r	r	r

Avoimen datan käyttötietojen kerääminen suoraan käyttöportaalien kautta käyttäjät tunnistamalla ei Etlan kyselyn perusteella onnistu tällä hetkellä useimpien keskeisten rekisteritietojen käytön osalta (kuvio 2.6 s. 17). Käyttäjien tunnistamisen ja käyttötietojen systemaattisen keräämisen esteet eivät ole merkittävässä määrin teknisiä. Suurin este käyttäjätietojen keräämiseen syntyy tiukasta avoimen datan käytön anonyymiuden periaatteen noudattamisesta. Keskeiset julkisten rekisteriaineistojen hallinnoijat pitävät usein avoimen datan hengen mukaisena, ettei tiedon käyttäjän tarvitse rekisteröityä avointa dataa tarjoavaan palveluun eikä tietoja käyttäviä henkilöitä tai yrityksiä pyritä tunnistamaan.

Avoimen datan vaikuttavuuden tutkimuksen näkökulmasta datan käyttäjien anonyymius hankaloittaa työtä. Avoimen datan käyttöä koskevia tietoja käyttöportaaleista voitaisiin kerätä lakisääteisten kyselytietojen tapaan luottamuksellisesti käsiteltävinä ja niin, että yksittäisiä käyttäjiä koskevat tiedot ovat salassa pidettäviä. Käytöstä saatavia tietoja voitaisiin tutkimuskäyttöä varten yhdistää esimerkiksi erilaisiin Tilastokeskuksen tietokantoihin luottamuksellisesti käsiteltävien, vain tutkimuskäyttöön luovutettavien aineistojen muodostamiseksi.

Suoraan käyttöportaaleista kerättävän datan käyttöä avoimen datan vaikuttavuuden tutkimuksessa harkittaessa on kuitenkin hyvä tiedostaa vaikuttavuustutkimukseen liittyvät muut rajoitteet. Kansalaisten osalta käyttäjien identifiointi mahdollistaisi avoimen datan käytön tarkemman tutkimuksen (esim. datan käyttäjien demografisten taustatekijöiden osalta), mutta tarjoaisi edelleen varsin niukasti mahdollisia taloudellisen vaikuttavuuden tutkimukseen odotettuja vaikutuksia kuvaavien mittarien puuttuessa. Kansalaiset ryhmänä ovat myös taloudellisen vaikuttavuusarvioinnin näkökulmasta hankala tutkimuskohde. Olemassa oleviin tilastoaineistoihin tai kerättäviin tietoihin ei sisälly suoraan kansalaistason odotettuja vaikutuksia kuvaavia mittareita. Datan hyödyntämisen taloudellisten vaikutusten arvioimisessa jouduttaisiin tukeutumaan pääosin kansalaisten omiin subjektiivisiin arvioihin hyödyistä. Esimerkiksi avoimen datan käytön ansiosta säästetyn ajan arvioiminen ei olisi realistinen tavoite.

Yritystasolla identifioitujen datan käyttötietojen soveltuvuus tutkimukseen on parempi kuin kansalaisten osalta. Avoimen datan taloudellisen vaikuttavuusanalyysin kohdistaminen yrityksiin olisi perusteltua myös siitä näkökulmasta, että yritykset muodostavat Etlan kyselyn perusteella merkittävimmän rekisteriaineistoja hyödyntävän käyttäjäryhmän. Tosin pelkkä tieto avoimen datan lataamisesta ei kerro sen käytöstä. Tämän lisäksi tarvittaisiin tarkentavaa kyselyihin perustuvaa tutkimusta siitä, mihin tarkoituksiin yrityksissä käytetään avointa dataa.

On huomattava, että avoin data muodostaa vain osan eri käyttäjäryhmien hyödyntämästä datasta. Datatalouden mekanismien tutkimuksen näkökulmasta olisi tärkeää, että datan käyttöä kuvastavaa tilastoaineistojen keräämistä kehitettäisiin. Näin pystyttäisiin muun muassa valottamaan laajemmin datan hyödyntämistä talouden eri sektoreilla sekä tutkimaan tarkemmalla tasolla datan käytön vaikuttavuutta.

3.2 Miten taloudellista vaikuttavuutta voidaan tutkia?

Avoimen datan vaikuttavuuden arvioimiseksi eivät riitä pelkät luvut siitä, miten toiminta tai taloudelliset mittarit ovat muuttuneet datan avaamisen jälkeen. Tällaisia ovat esimerkiksi deskriptiiviset kuvaukset siitä, kuinka paljon avointa dataa hyödyntäneet yritykset ovat kasvaneet tai luoneet uusia työpaikkoja. Ei ole mitenkään mahdollistakaan tietää sitä, kuinka paljon avoimen datan hyödyntäjäyrityksissä olisi syntynyt uusia työpaikkoja tai kuinka paljon ne olisivat kasvaneet tilanteessa, jossa data olisi ollut suljettua. Pelkät deskriptiiviset luvut ilman huo-

llestä tilastollista analyysia eivät sen takia anna aitoa kuvaa vaikuttavuudesta eivätkä täytä tutkimuksellisia vaikuttavuusarvioinnin kriteereitä.

Avoimen datan vaikuttavuuden tutkimuksen näkökulmasta ihanteellinen tutkimusasetelma olisi sellainen, jossa voitaisiin verrata datan avaamisen jälkeistä kehitystä (esim. yritysten uusien dataan perustuvien innovaatioiden määrä) vastaavaan kehitykseen datan ollessa suljettua. Sitä, mitä olisi tapahtunut datan käytölle ja vaikuttavuuden mittaamiseen käytetyille mittareille (esim. yrityksen kasvu), ei voida kuitenkaan käytännössä havaita. Ihanteellista tutkimusasetelmaa voidaan pyrkiä jäljittelemään käyttämällä ekonometrisiä menetelmiä.

Tutkittaessa avoimen datan vaikutuksia yritystasolla on yksi vaihtoehto verrata avointa dataa hyödyntäneiden yritysten kehitystä mahdollisimman samanlaisten, avointa dataa hyödyntämättömien yritysten kehitykseen. Tätä kutsutaan vertaistamiseksi. Toinen ekonometrisen työkalupakin tarjoama menetelmä on ns. instrumenttimuuttujamenetelmä. Tämän menetelmän soveltaminen kuitenkin edellyttää sitä, että saatavilla on jokin muuttuja eli instrumentti, joka vaikuttaa yrityksen valintaan käyttää avointa dataa, mutta ei suoraan yrityksen kehitystä mittaavaan tekijään. Kolmas vaihtoehto, ns. erotukset erotuksissa -menetelmä, vertaa avointa dataa käyttäneiden yritysten kehitystä käytetyllä mittarilla ennen ja jälkeen avoimen datan hyödyntämistä avointa dataa hyödyntämättömien yritysten tulevien erotukseen ko. mittarilla samana ajanjaksona. Erotukset erotuksissa -menetelmä edellyttää tietoa yritysten avoimen tiedon käytöstä ja niiden toiminnasta tai taloudellisesta kehityksestä useamman vuoden ajalta.

Tällä hetkellä olemassa olevat aineistot avointa dataa hyödyntävistä yrityksistä ovat varsin suppeita, kyselyaineistoihin perustuvia otoksia. Kehittyneiden ekonometristen vaikuttavuusarvointiin soveltuvien menetelmien käyttäminen nykytilanteessa on hankalaa. Esimerkiksi avoimen datan käytön vaikutuksia uusien tuotteiden ja palveluiden kehittämiseen ei voida tutkia olemassa olevan CIS-innovaatiokyselyaineiston avulla vertaistamismenetelmää käyttäen: avointa dataa hyödyntäville innovaatiotoimintaa harjoittaville yrityksille ei löydy riittävän suurta määrää samanlaisia avointa dataa käyttämättömiä ja innovaatiotoimintaa harjoittavia yrityksiä verrokeiksi.

Instrumenttimuuttujamenetelmän hyödyntäminen on myös hankalaa, koska aineistossa ei ole sellaisia muuttujia, joiden voitaisiin olettaa vaikuttavan vain yrityksen todennäköisyyteen käyttää avointa dataa, mutta ei niiden innovaatioiden tuotantoon. Tietoja yritysten avoimen datan hyödyntämisestä ei ole myöskään saatavissa riittävän pitkältä ajanjaksolta erotukset erotuksissa -menetelmän käyttämiseksi. Sama pätee Etlan vuonna 2011 keräämään kyselyaineistoon yritysten avoimen datan hyödyntämisestä.

"Tiukasti" vaikuttavuutta mittaavien menetelmien sijaan olemme tämän takia käyttäneet sellaisia kehittyneitä ekonometrisiä mallintamismenetelmiä, joiden avulla voidaan analysoida avointa dataa innovaatiotoiminnassaan hyödyntävien yritysten kehitystä verrattuna avointa dataa käyttämättömien yritysten kehitykseen tietyillä taloudellisilla mittareilla. Mittareina käytetään i) yritysten todennäköisyyttä tuottaa uusia tuotteita ja palveluita ja ii) yritysten kasvua liikevaihdolla mitattuna. Käytetyt estimointimenetelmät kontrolloivat useita taustatekijöitä, joiden taloustieteellinen kirjallisuus on havainnut liittyvän innovaatioiden tuottamiseen ja kasvuun yrityksissä. Tällaisia tekijöitä ovat muun muassa yrityksen ikä, koko, toimiala ja sijainti. Estimointimme ottavat siis huomioon joukon muita innovaatioiden tuottamista ja kasvua selittäviä tekijöitä avoimen tiedon käytön lisäksi. Täten esimerkiksi toimialakohtaiset erot innovaatioiden tuottamisen todennäköisyydessä huomioidaan ja erotetaan avoimen datan käytön yhteydestä innovaatioiden tuottamisen todennäköisyyteen.

3.3 Aineistoanalyysin tulokset

3.3.1 Avoimesta datasta innovaatioita?

Tässä osiossa tarkastellaan avoimen datan käyttöä ja sen vaikutuksia yritysten tuotteisiin liittyvässä innovaatiotoiminnassa. Innovaatiotoiminta luo edellytyksiä yrityksen kilpailukykyyn ja markkina-aseman kohentumiselle. Tuotteisiin liittyvillä innovaatioilla yritys voi tuoda markkinoille täysin uudenlaisia tuotteita tai lisätä nykyisiin tuotteisiin uusia tai olennaisesti parannettuja ominaisuuksia. Nämä toimet voivat kasvattaa yrityksen liikevaihtoa ja voiton-tekemahdollisuuksia. Aineistona tarkastelussa käytetään Tilastokeskuksen vuoden 2014 innovaatiotutkimusta, jossa on kerätty tietoja yritysten innovaatiotoiminnasta koskien vuosia 2012–2014.

Avointa dataa innovaatiotoiminnassa hyödyntävien yritysten ominaispiirteitä

Avoimen datan käyttö tuotteisiin liittyvässä innovaatiotoiminnassa on yleisintä yrityskokoluokittain tarkasteltuna suuremmissa yrityksissä: vähintään 250 henkeä työllistävistä yrityksistä yli viidennes on hyödyntänyt avointa dataa innovaatiotoiminnassaan, kun vastaava osuus pienissä, 10–49 henkeä työllistävissä yrityksissä on vain runsas kymmenes (innovaatiotutkimus ei kata alle 10 henkeä työllistäviä yrityksiä). Yrityksen iän suhteen emme havaitse suuria eroja avoimen datan hyödyntämisen osalta, mutta suhteellisesti yleisintä se kuitenkin on nuorissa, alle viisi vuotta vanhoissa yrityksissä. Toimialoittain tarkasteltuna avointa dataa hyödynnetään tuotteisiin liittyvässä innovaatiotoiminnassa suhteellisesti yleisimmin tieteellisessä ja teknisessä toiminnassa, rahoitus- ja vakuutusallalla sekä energia- ja vesihuollossa ja ympäristönhoidossa.

Digitalisaatiolla, ts. tavaroiden, palveluiden, niiden tuotannon tai jakelun siirtymisellä sähköiseen muotoon, on huomattava yhteys avoimen datan käytön kanssa innovaatiotoiminnassa. Yli viidennes sellaisista yrityksistä, joissa digitaalisten tavaroiden ja/tai palveluiden merkitys liiketoiminnalle on kohtalainen tai suuri, on käyttänyt avointa dataa innovaatiotoiminnassaan. Vastaava osuus on alle 3 % sellaisten yritysten joukossa, joilla digitaalisilla tuotteilla ei ole merkitystä liiketoiminnalle. Avoimen datan ja massadatan ("big data") käytöllä on myös vahva korrelaatio. 42 % niistä yrityksistä, joilla massadatan käyttö innovaatiotoiminnassa on merkittävää, on hyödyntänyt avointa dataa innovaatiotoiminnassaan vuosina 2012–2014. Vastaa- vasti vain 5 % yrityksistä, joilla massadatan käytöllä innovaatiotoiminnassa ei ole merkitystä, on käyttänyt avointa dataa tuoteinnovoinnissa.

Avoimen datan hyödyntäminen tuoteinnovaatioiden tuottamisessa on yhteydessä innovaatioyhteistyöhön julkisen sektorin kanssa. Yli neljännes yrityksistä, joilla on innovaatioyhteistyötä julkisen sektorin kanssa, hyödyntää avointa dataa tuoteinnovoinnissa. Vastaava osuus on vain noin 10 % niiden yritysten joukossa, joilla ei ole ollut innovaatioyhteistyötä julkisen sektorin kanssa. Avointa dataa käyttävillä on myös muilta osin keskimäärin useampia yhteistyöta- hoja, kuten muita saman alan yrityksiä tai tutkimuslaitoksia, kuin sitä käyttämättömillä. Niiden innovaatiotoiminta on siis luonteeltaan verkostomaisempaa kuin muilla yrityksillä.

Avoimen datan käytön yhteys markkinoiden kannalta uusiin tuoteinnovaatioihin

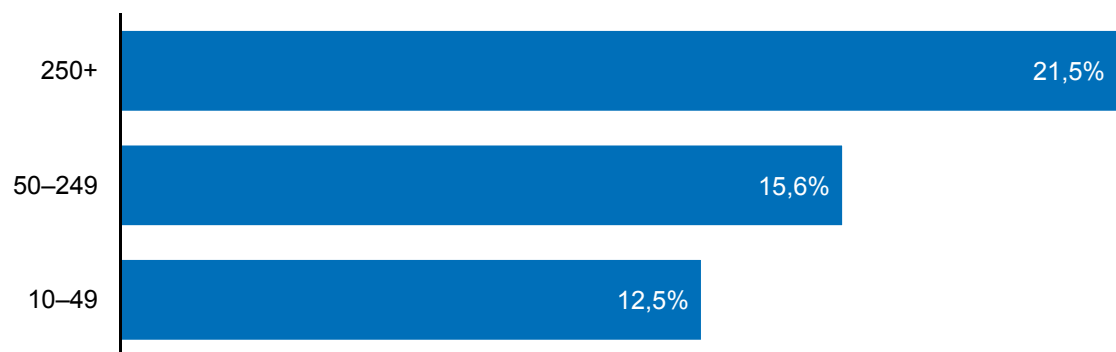
Tuoteinnovaatiot voidaan jakaa kahteen tyyppiin. Innovaatiot voivat olla yksinkertaisimmillaan uusia ainoastaan sen tehneen yrityksen kannalta, jolloin vastaavia tuotteita on ollut jo saata- vissa yrityksen kilpailijoilta. Näitä merkittävämpiä ovat sellaiset yrityksen tekemät innovaatiot,

joita se on tuonut markkinoilleen ennen kilpailijoitaan. Tässä keskitytään nimenomaan näiden ”radikaalien” tuoteinnovaatioiden tarkasteluun.

Avointa dataa innovaatiotoiminnassaan hyödyntävät yritykset tekevät markkinoiden kannalta uusia tavara- ja/tai palveluinnovaatioita suhteellisesti huomattavasti useammin kuin muut yritykset. Jopa yli puolet avointa dataa käyttäneistä innovaatiotoimintaa harjoittaneista yrityksistä on tehnyt vuosina 2012–2014 radikaaleja tuoteinnovaatioita. Vastaava osuus on vain runsas kolmannes muiden innovaatiotoimintaa harjoittavien yritysten joukossa. Osuus on erityisen korkea sellaisten innovaatiotoimintaa harjoittavien yritysten keskuudessa, jotka käyttävät avointa dataa ja joissa samalla digitaalisten tuotteiden merkitys liiketoiminnassa on merkittävää. Tällaisista yrityksistä jopa kolme viidestä on tehnyt markkinoidensa kannalta uusia tuoteinnovaatioita.

Jaoteltaessa erikseen palveluihin ja tavaroihin liittyviä radikaaleja tuoteinnovaatioita havaitaan, että avoimen datan käyttäjät tuottavat näitä molempia tyyppisiä huomattavasti todennäköisemmin kuin muut innovaatiotoimintaa harjoittavat yritykset. Ero muihin yrityksiin on

Kuvio 3.1 Osuus innovaatiotoimintaa harjoittavista yrityksistä, joissa julkisen sektorina avoimen datan käyttö tuoteinnovaatioissa on ollut kohtalaista tai suurta vuosina 2012–2014 yritysten Suomessa työskentelevän henkilöstömäärän mukaan jaoteltuna



Aineistolähde: Tilastokeskuksen innovaatiotutkimus 2014 (Raportoituna painottamattomat otoskeskiarvot).

Kuvio 3.2 Osuus innovaatiotoimintaa harjoittavista yrityksistä, joissa julkisen sektorina avoimen datan käyttö tuoteinnovaatioissa on ollut kohtalaista tai suurta vuosina 2012–2014 yritysten iän mukaan jaoteltuna

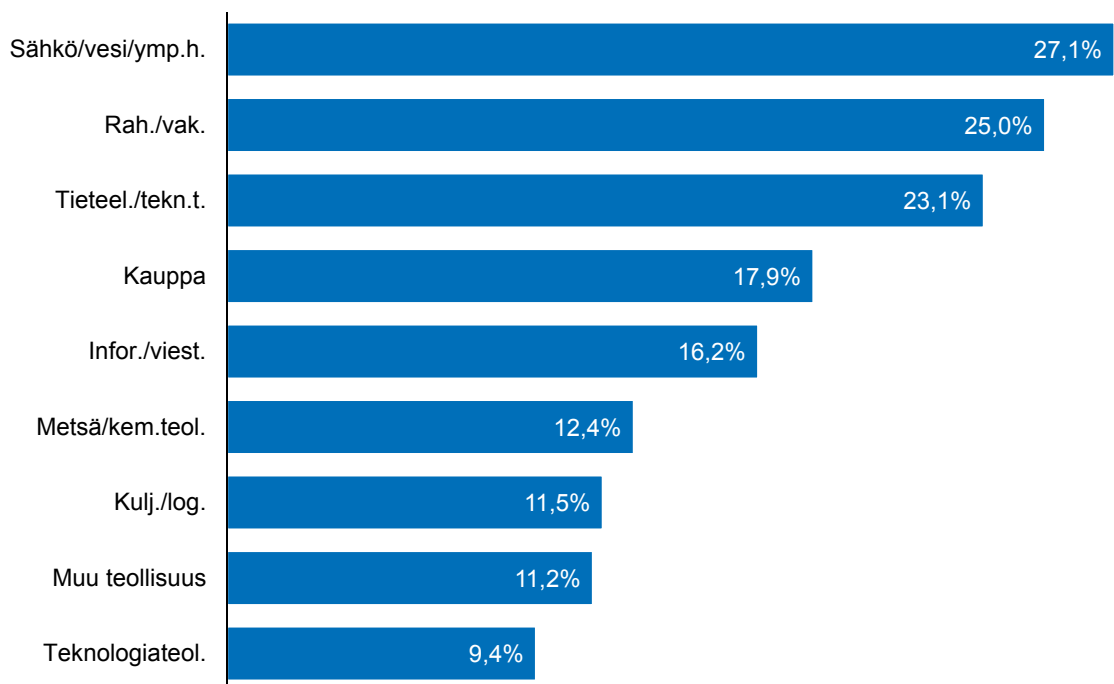


Aineistolähde: Tilastokeskuksen innovaatiotutkimus 2014 (Raportoituna painottamattomat otoskeskiarvot).

kuitenkin merkittävien palveluihin liittyvissä innovaatioissa, joita avointa dataa hyödyntävät yritykset tekevät noin kaksi kertaa todennäköisemmin kuin muut innovaatiotoimintaa harjoittavat yritykset.

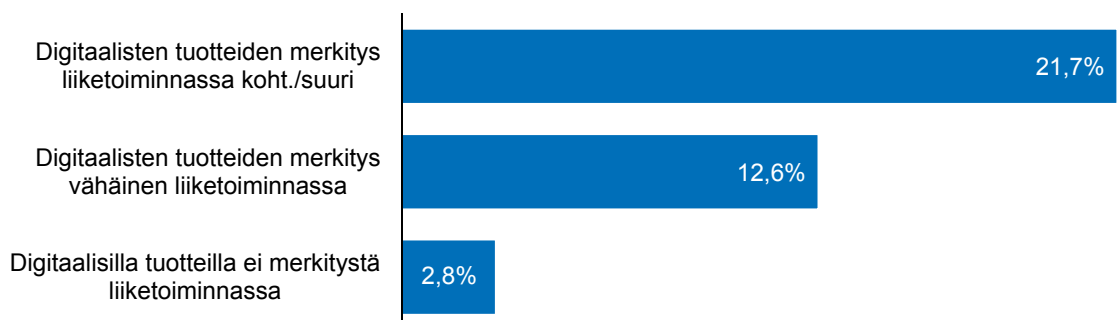
Avoimen datan käyttäjiä erikseen omana ryhmänä tarkasteltaessa tuoteinnovaatiot liittyvät enemmän palveluihin kuin tavaroihin. Kaikista avoimen datan käyttäjistä yli 40 % on tehnyt palveluihin liittyviä markkinoiden kannalta uusia tuoteinnovaatioita. Kyseinen osuus on jopa

Kuvio 3.3 Osuus innovaatiotoimintaa harjoittavista yrityksistä, joissa julkisen sektorina avoimen datan käyttö tuoteinnovaatioissa on ollut kohtalaista tai suurta vuosina 2012–2014 yritysten toimialan mukaan jaoteltuna



Aineistolähde: Tilastokeskuksen innovaatiotutkimus 2014 (Raportoituna painottamattomat otoskeskiarvot).

Kuvio 3.4 Osuus innovaatiotoimintaa harjoittavista yrityksistä, joissa julkisen sektorina avoimen datan käyttö tuoteinnovaatioissa on ollut kohtalaista tai suurta vuosina 2012–2014 jaoteltuna sen mukaan, kuinka merkittävää digitaaliset tuotteet ovat yrityksen liiketoiminnalle

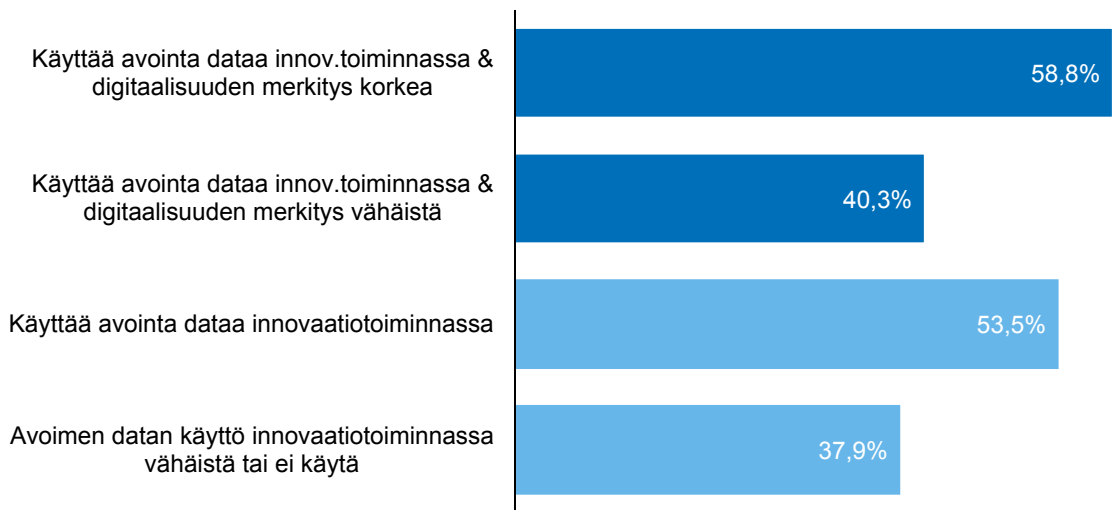


Aineistolähde: Tilastokeskuksen innovaatiotutkimus 2014 (Raportoituna painottamattomat otoskeskiarvot).

lähes puolet sellaisista avoimen datan käyttäjistä, joiden liiketoiminnassa digitaalisten tuotteiden merkitys on huomattava. Vastaavasti tavaroihin liittyviä innovaatioita avoimen datan käyttäjistä on tehnyt noin 35 % ja 38 % sellaisista avoimen datan käyttäjistä, joilla digitaalisten tuotteiden merkitys liiketoiminnassa on huomattava.

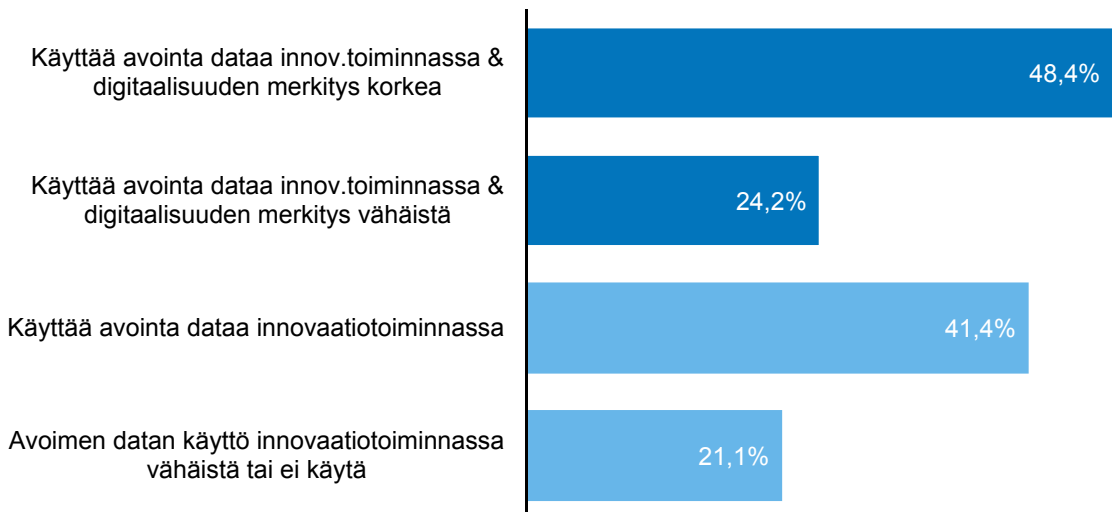
Aineistolähteessämme on jaoteltu avoimen datan käyttö innovaatiotoiminnan osalta myös yksityiskohtaisemmin tuotteiden kehittämiseen ja toisaalta niiden parantamiseen liittyvään

Kuvio 3.5 Osuus yrityksistä, jotka toivat markkinoille vuosina 2012–2014 uusia tai olennaisesti parannettuja tavaroita tai palveluita jaoteltuna avoimen datan käytön ja digitaalisten tuotteiden liiketoimintamerkityksen suhteen



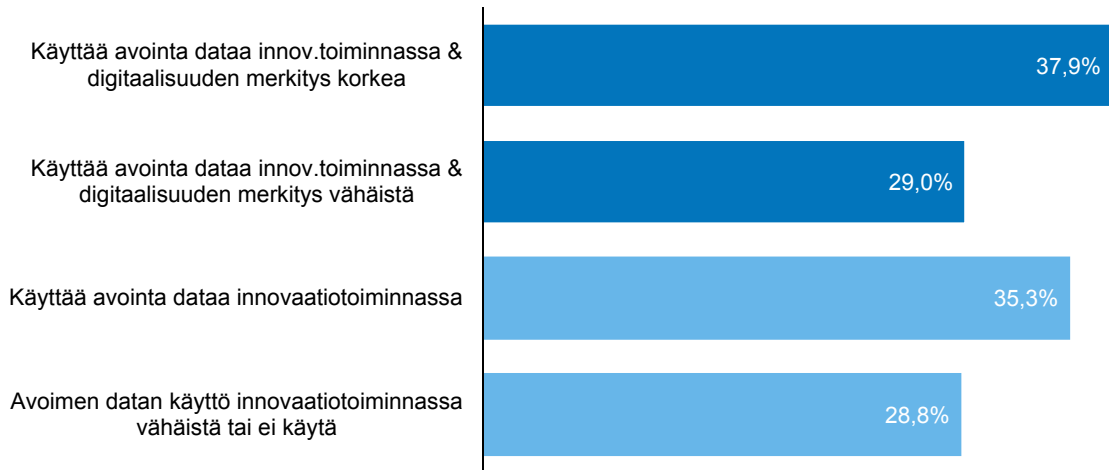
Aineistolähde: Tilastokeskuksen innovaatiotutkimus 2014 (Raportoituna painottamattomat otoskeskiarvot).

Kuvio 3.6 Osuus yrityksistä, jotka toivat markkinoille vuosina 2012–2014 uusia tai olennaisesti parannettuja palveluita jaoteltuna avoimen datan käytön ja digitaalisten tuotteiden liiketoimintamerkityksen suhteen



Aineistolähde: Tilastokeskuksen innovaatiotutkimus 2014 (Raportoituna painottamattomat otoskeskiarvot).

Kuvio 3.7 Osuus yrityksistä, jotka toivat markkinoille vuosina 2012–2014 uusia tai olennaisesti parannettuja tavaroita jaoteltuna avoimen datan käytön ja digitaalisten tuotteiden liiketoimintamerkityksen suhteen



Aineistolähde: Tilastokeskuksen innovaatiotutkimus 2014 (Raportoituna painottamattomat otoskeskiarvot).

toimintaan. Tässä suhteessa emme havainneet merkittäviä poikkeamia osuuksissa, vaan avointa dataa näytetään hyödynnettävän varsin yhtäläisesti sekä uusien tuotteiden kehittämisessä että nykyisten tuotteiden parantamisessa.

Halusimme lopuksi testata ekonometrisin menetelmin, säilyykö avoimen datan käyttö merkitsevänä tekijänä, kun huomioimme yritysten muita taustatietoja. Analyysissa ”vakioimme” yritysten koon, iän, toimialan ja maantieteellisen sijainnin vaikutukset selitettäessä markkinoiden kannalta uusien tuoteinnovaatioiden tekemisen todennäköisyyttä. Tuloksissa avoimen datan käyttö oli tilastollisesti merkitsevä tekijä muut yllä mainitut taustatekijät huomioituina sekä kaikkien tuoteinnovaatioiden osalta että erikseen palveluihin ja tavaroihin liittyvien innovaatioiden todennäköisyyttä selitettäessä. Samoin avoimen datan käyttö säilyi tilastollisesti merkitsevänä selittäjänä tuoteinnovaatioille – muut taustatekijät huomioituna – sellaisissa yrityksissä, joissa digitaalisten tuotteiden merkitys liiketoiminnalle on huomattavaa. Lisäksi havaitsimme, että avoimen datan käyttö on vahvasti korreloinut massadatan käytön kanssa, ja löysimme samanlaisen yhteyden massadatan käytön ja innovaatioiden tuottamisen todennäköisyyden välillä.

Johtopäätöksemme on, että dataa innovaatiotoiminnassaan hyödyntävät yritykset tekevät muita yrityksiä huomattavasti todennäköisimmin markkinoiden kannalta uusia tuoteinnovaatioita. Erityisen todennäköistä tämä on niiden dataa hyödyntävien yritysten keskuudessa, joissa digitaalisten tuotteiden merkitys liiketoiminnalle on huomattavaa.

3.3.2 Luoko datan käyttö innovaatioissa kasvua?

Tässä osiossa on hyödynnetty Etlan syksyllä 2011 yritys­kyselyn avulla koottua aineistoa, johon on yhdistetty Asiakastiedon tilinpäätöstietoja vuosilta 2010–2014 kyseisistä yrityksistä. Kyselytutkimuksen tarkoituksena oli kartoittaa suomalaisten yritysten datan käyttöä, ja erityisen julkisen sektorin tietojen hyödyntämistä, uusien palveluiden ja tuotteiden kehittämisessä.

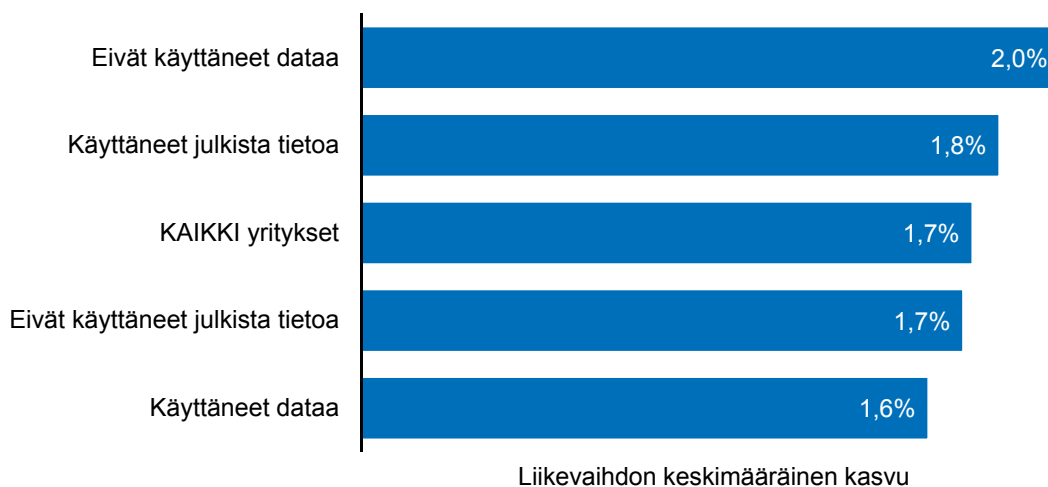
Kysely kohdennettiin laajalti eri toimialojen yrityksille. Vastaus saatiin 531 yritykseltä, joista 80 prosenttia toimi palvelusektorilla. Suuri osa vastanneista yrityksistä, 93 prosenttia, oli joko alle 10 henkeä työllistäviä mikroyrityksiä tai pieniä 10–49 henkeä työllistäviä yrityksiä.

Kyselytutkimuksessa kysyttiin oliko yritys käyttänyt viimeisen kolmen vuoden aikana (syksyn 2011 tilanne) uusissa tuotteissaan ja/tai palveluissaan erilaisia tietotyyppejä: paikkatietoa, säätietoa, väestötietoa, liikennetietoa, yritystietoa ja/tai muita tietoja. Vastauksessa eroteltiin myös tietotyyppien tiedonhankintakanavat, ts. i) tuottiko yritys tiedon itse, ii) hankkiko tai ostiko se tiedon julkiselta sektorilta, iii) hankkiko tai ostiko se tiedon tietoa tuottavalta yritykseltä ja/tai iv) hankkiko se tiedon ilmaispalvelusta. Aineistosta voidaan erottaa julkisen tiedon hyödyntäjät, mutta aineiston perusteella ei ole mahdollista erottaa avoimen datan käyttäjiä julkisen sektorin avaamattomien tietojen käyttäjistä.

Estimointitulokset viittaavat siihen, etteivät julkisia tietoja tai dataa ylipäätään uusien tuotteiden ja/tai palveluiden kehittämisessä käyttäneet yritykset ole kasvaneet muita yrityksiä nopeammin liikevaihdon mitattuna. Aineistoanalyysi, jossa kontrolloidaan samanaikaisesti useita potentiaalisia liikevaihdon kasvuun vaikuttavia tekijöitä, on hyvin samansuuntainen kuviossa 3.8 esitettyjen deskriptiivisten lukujen kanssa. Kaikkien yritysten ja julkisen sektorin tai muuta dataa hyödyntäneiden ja ei-hyödyntäneiden yritysten vuosittainen keskimääräinen kasvuvauhti liikevaihdon mitattuna on lähellä toisiaan, 1,6–2,0 prosenttia.

Datan hyödyntämisen yhteyttä yrityksen kasvuun arvioitiin myös toimialakohtaisesti. Useilla toimialoilla havaintojen määrä oli riippuvuussuhteen estimointiin liian vähäinen. Informaatio- ja viestintätoimialan⁴ yrityksistä ne, jotka olivat hyödyntäneet dataa uusien tuotteiden ja/tai palveluiden kehittämisessä kasvoivat vuosina 2012–2014 keskimäärin nopeammin kuin muut yritykset. Estimointitulokset osoittavat, että syksyllä 2011 tai sitä edeltävänä kolmena vuonna dataa innovaatioiden kehittämisessä käyttäneiden informaatio- ja viestintätoimialan yritysten

Kuvio 3.8 Kaikki yritykset: datan käyttö uusissa tuotteissa ja/tai palveluissa ja liikevaihdon keskimääräinen kasvu 2012–2014



Aineistolähde: Etlan yritys­kysely 2011 & Asiakastiedon tilinpäätöstiedot.

⁴ "Tähän pääluokkaan kuuluvat tieto- ja kulttuurituotteiden valmistus ja jakelu, näiden tuotteiden sekä tiedon ja viestien siirto ja jakelu, tietotekninen palvelutoiminta sekä tietojen käsittely ja muu tietopalvelutoiminta. Mainosajan tai -tilan myyminen suoraan joukkotiedotusvälineen tai mainospaikan omistajan (esim. kustantaja) toimesta tms. luokitellaan myös tähän toimialaan." <http://www.stat.fi/meta/luokitukset/toimiala/001-2008/fj.html>

liikevaihto kasvoi vuosina 2012–2014 keskimäärin yli 17 prosenttia enemmän kuin saman toimialan dataa innovaatioissa hyödyntämättömien yritysten liikevaihto. Julkista tietoa innovaatioiden kehittämisessä käyttäneiden toimialan yritysten liikevaihdon kasvu ei poikennut tilastollisesti merkitsevästi niiden yritysten kasvusta, jotka eivät käyttäneet julkista tietoa.

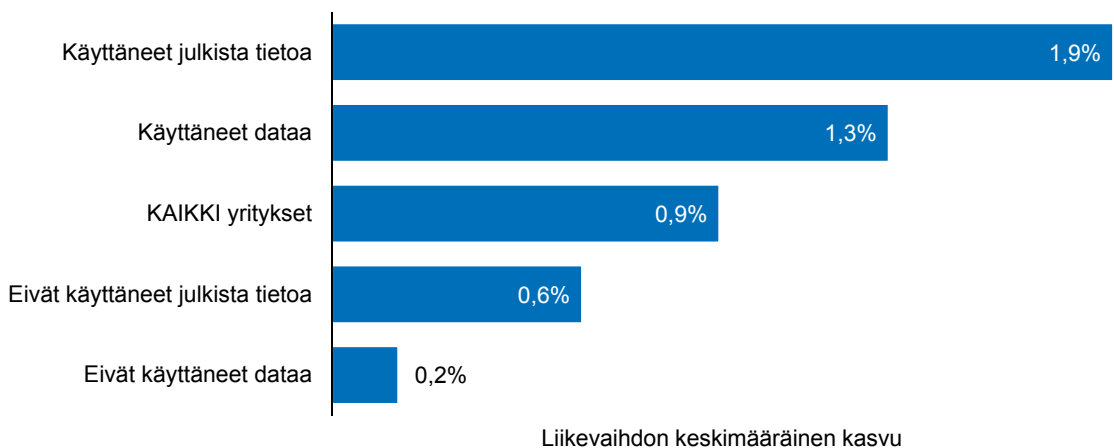
Yritysten kasvu työntekijämäärällä mitattuna ei myöskään poikkeaa tilastollisesti merkittävästi julkista tietoa tai muuta dataa innovaatioiden kehittämisessä hyödyntäneiden yritysten välillä. Estimointitulokset ovat samansuuntaisia kuin liikevaihdon kasvunkin kohdalla informaatio- ja viestintätoimialan yritysten kohdalla. Kyseisen toimialan dataa hyödyntäneiden yritysten työntekijämäärä kasvoi vuosina 2012–2014 keskimäärin 18 prosenttia enemmän kuin muiden, dataa hyödyntämättömien saman toimialan yritysten työntekijämäärä. Sen sijaan julkisen sektorin tietojen hyödyntäminen ei tälläkään toimialalla liittynyt tilastollisesti merkittävästi työntekijämäärän kasvuun.

Kuvio 3.9 Informaatio- ja viestintäsektorin yritykset: datan käyttö uusissa tuotteissa ja/tai palveluissa ja liikevaihdon keskimääräinen kasvu 2012–2014



Aineistolähde: Etlan yritys­kysely 2011 & Asiakastiedon tilinpäätöstiedot.

Kuvio 3.10 Yritysten datan käyttö uusissa tuotteissa ja/tai palveluissa ja työntekijöiden määrän keskimääräinen kasvu 2012–2014



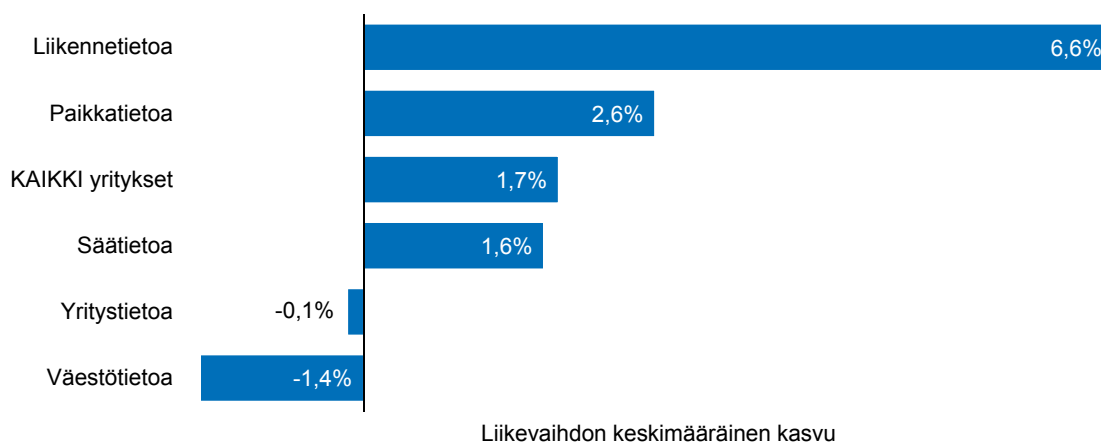
Aineistolähde: Etlan yritys­kysely 2011 & Asiakastiedon tilinpäätöstiedot.

Analysoimme myös poikkeako eri tietotyyppejä uusien tuotteiden ja palveluiden kehittämisessä käyttävien yritysten kasvu muiden yritysten kasvusta. Aineistoanalyysi viittaa siihen, että yksittäisistä tietotyypeistä vain liikennetiedoilla on positiivinen ja tilastollisesti merkitsevä yhteys liikevaihdon kasvuun. Myös deskriptiiviset havainnot osoittavat, että liikennetietoa innovaatioiden kehittämisessä käyttäneiden yritysten liikevaihdon kasvu on ollut huomattavasti suurempaa kuin yritysten kasvu keskimäärin. Liikennetietojen hyödyntämisen suureen taloudelliseen merkitykseen viittaa myös Trafín tietoluovutusten ja avoimen datan latausten suuri määrä; Trafín vuoden 2015 tietotilinpäätöksen mukaan sen tietoja hyödynnettiin kyseisenä vuonna enemmän kuin koskaan (kts. luku 4.2.3).

Paikkatietoa hyödyntäneiden yritysten liikevaihto kasvoi myös keskimääräistä nopeammin, mutta ero muihin yrityksiin ei ollut ekonometrisen analyysin perusteella tilastollisesti merkitävä. Eri tietotyyppien käyttö ei liittynyt tilastollisesti merkittävästi yritysten työntekijämäärän muutoksiin. Myöskään innovaatioissa käytettyjen julkisten tietojen osalta eri tietotyypeillä ei ollut tilastollisesti merkittävää yhteyttä yritysten kasvuun.

Aineistoanalyysi viittaa siihen, että dataa uusien palveluiden ja/tai tuotteiden kehittämisessä hyödyntäneet yritykset ovat kasvaneet muita yrityksiä nopeammin joillakin tietointensiivisillä sektoreilla⁵. Tämä tukee hyvin näkemystä siitä, että eri yritysten potentiaali hyödyntää dataa liittyy läheisesti yritysten toimintojen tietointensiivisyyteen. Informaatio- ja viestintäsektorin datan hyödyntäjien liikevaihto ja työntekijämäärä ovat selvästi kasvaneet muita yrityksiä nopeammin. Aineiston informaatio- ja viestintäalan yrityksistä liki 80 prosenttia toimi ”Ohjelmistot, konsultointi ja siihen liittyvä toiminta” -alatoimialalla, ja suurimman joukon näistä muodostivat ohjelmistoja suunnittelevat ja valmistavat yritykset. Toisin sanoen, analyysin tulokset viittaavat vahvasti siihen, että dataa uusien ohjelmistoratkaisujen suunnittelussa käyttäneet yritykset ovat kasvaneet muita voimakkaammin. On huomattava, että monien toimialojen osalta havaintojen määrä ei riitä johtopäätösten tekemiseen datan hyödyntämisen toimialoitaisista vaikutuksista.

Kuvio 3.11 Liikevaihdon keskimääräinen kasvu 2012–2014 yrityksissä, jotka käyttäneet uusissa tuotteissaan ja/tai palveluissaan



Aineistolähde: Etlan yritys­kysely 2011 & Asiakastiedon tilinpäätöstiedot.

⁵ McKinsey Global Institutin (2011) raportti arvioi tietointensiivimpien toimialojen käsittävän rahoituspalvelut, informaatio- ja viestintäsektorin sekä sähkö-, kaasu- ja lämpöhuollon. Näiden toimialojen yrityksillä on hallussaan eniten dataa organisaatiota kohden.

Julkista tietoa uusien palveluiden ja/tai tuotteiden kehittämisessä vuoden 2011 syksyllä tai sitä edeltävänä kolmena vuonna käyttäneet yritykset eivät ole kasvaneet muita yrityksiä nopeammin. On huomattava, että monet tällä hetkellä avoimena tarjolla olevista julkisista tietovarannoista on avattu vasta vuoden 2011 jälkeen: Maanmittauslaitoksen kaikki digitaaliset maastoaineistot avattiin toukokuussa 2012, Ilmatieteen laitoksen säätiedot avattiin vuotta myöhemmin ja Liikenteen turvallisuusvirasto Trafi aloitti tietojensa avaamisen kesäkuussa 2014. Tämän takia edellä esitetyistä tutkimuslöydöksistä ei voida vetää laajakantoisia johtopäätöksiä julkisen sektorin avoimen datan ja yritysten kasvun välisestä yhteydestä.

Aiempaa tutkimusta datan hyödyntämisen ja yritysten kasvun välisestä yhteydestä on varsin vähän. Kosken (2011) aineistoanalyysi 15 eri maan yritysten joukossa antaa viitettä siitä, että paikkatietojen avaaminen ja rajakustannushinnoitteluun siirtyminen olivat yhteydessä nopeampaan kasvuun ”tekninen palvelu” -toimialan yritysten joukossa. Niissä maissa, joissa paikkatietoja tarjottiin maksutta tai rajakustannushintaan, toimialan yritysten liikevaihto kasvoi vuositasolla keskimäärin 15 prosenttia nopeammin kuin saman toimialan yritysten liikevaihto muissa maissa. Paikkatietojen avoimuuden ja yritysten kasvun välillä ei löytynyt kuitenkaan yhteyttä suurten yritysten osalta. Tässäkin luvussa raportoidut tulokset datan hyödyntämisen ja yritysten kasvun välisestä yhteydestä koskevat pääosin pieniä yrityksiä aineiston muodostuessa suurelta osin alle 50 henkeä työllistävistä yrityksistä.

4. AVOIMEN DATAN HYÖDYNTÄMINEN

Mika Honkanen – Teemu Ropponen

Suomi on maailman johtavia avoimen datan maita. Avoin data on merkittävä digitaalisen yhteiskunnan rakennusaine. Julkishallinnon tulisi tavoitella kokonaisvaltaista tiedonhallinnan ohjausta, joka kattaa ydintiedot, metatiedot, asiatiedot ja avoimen tiedon. Selvityksen tärkeimmät löydökset liittyvät organisaatioiden tiedonhallintaan ja tietojohdamiseen, käyttäjälähtöiseen toimintaan, avoimen datan laadun kehittämiseen ja lainsäädännön epätarkkuuteen. Suuri osa tutkittavista rekistereistä sisälsi paljon henkilötietoja. Myös henkilötietoja voitaisiin hyödyntää laajemmin, mm. omadatan ja anonymisoinnin avulla, kun huomioidaan oikeudelliset ja tekniset haasteet.

4.1 Tietojen kerääminen

Tutkimuksen kolmantena tavoitteena oli selvittää (tutkimuskysymys), millaisin **politiikkatoimin** ja millaisilla **toimintatavoilla** tai **rekistereiden rakenteellisilla muutoksilla** voidaan tehokkaimmin edistää julkisen hallinnon tietovarantojen hyödyntämistä laajasti niin liiketoiminnassa kuin yhteiskunnassa yleisesti.

Tutkimustyökaluksi valittiin PESTL-analyysi, joka on kymmenen eniten käytetyimmän (Bensoussan & Fleisher, 2008, s. 11) analyysi-viitekehityksen joukossa. Analyysimenetelmä ohjaa tarkastelemaan tutkimusongelmaa monipuolisesti, jotta analysoitavaan toimintaympäristöön merkittävästi vaikuttavia asioita ei jätetä huomioimatta. Menetelmä suuntaa huomioita lyhyen aikavälin ja osien optimoinnin sijasta kokonaisuuteen ja pidemmän aikavälin tavoitteisiin. PESTL-analyysissa tutkimuskysymystä tarkastellaan viidestä näkökulmasta, jotka eivät ole tarkkaan rajattuja. Niinpä osa huomioista voi kuulua useampaan näkökulmaan:

1. Poliittinen (Political, P)
2. Taloudellinen (Economic, E)
3. Sosiaalis-kulttuurinen (Socio-cultural, S)
4. Teknologinen (Technological, T)
5. Oikeudellinen (Legal, L)

Analyysin empiirinen aineisto kerättiin teemahaastatteluissa, työpajoissa ja julkisesti saatavissa olevasta materiaalista (esimerkiksi internet-sivut, avoimen datan palvelut, tietotilinpäätökset ja tulossopimukset) avulla. Tutkimuksen aikana kerättiin verkossa olleeseen raportin luonnokseen tietoa ja näkemyksiä eri sidosryhmiltä avoimella kommentoinnilla, johon osallistui noin 15 henkilöä.

Tutkimuksessa kerättiin empiiristä tietoa teemahaastattelujen avulla 12 organisaatiolta, (ks. luku 2.2), joista kustakin haastatteluihin osallistui 1–4 henkilöä. Teemahaastattelu ei etene tarkkojen valmiiksi muotoiltujen kysymysten avulla, vaan keskusteluna väljemmin kohdenne-tuista aihealueista niin, että haastateltavien vapaalle puheelle annettiin tilaa (liite 3). Haastattelut tehtiin 13.10.2016–2.12.2016. Osa haastateltavista oli eri henkilöitä kuin Etlan kyselyyn aiemmin vastanneet henkilöt. Haastattelut äänitettiin ja litteroitiin. Haastattelukommenteista on poistettu yksilöivä tieto, josta henkilö voitaisiin tunnistaa. Lisäksi osa haastateltavista on saattanut toivoa, että heitä ei ilman lupaa siteerata suoraan.

Osana avointa ja osallistavaa toimintatapaa järjestettiin kaksi kolmetuntista työpajaa Helsingissä. Työpajoihin lähetettiin avoin kutsu, jota jaettiin mm. haastateltujen organisaatioiden kautta avoimen datan asiantuntijoille ja laajemmin avoimen datan yhteisölle mm. sosiaalisen median kanavien kautta. Työpajoihin osallistui henkilöitä erityisesti yrityksistä ja julkishallinnosta, jonkin verran yliopistoista ja kansalaisjärjestöistä sekä muutama yksityishenkilö.

Ensimmäisessä työpajassa (20.12.2016, 19 osallistujaa) kartoitettiin avoimen datan haasteita ja kehitettiin niihin mahdollisia ratkaisuaihioita. Avaamisen haasteet ja esitetyt ratkaisuaihiot priorisoitiin äänestämällä työpajassa, mikä ohjasi myöhempää työtä. Työstettyjä näkökulmia olivat juridiikka, vuorovaikutus datan tuottajien ja hyödyntäjien välillä, tiedon tuottajien toimintatavat, datan saatavuus ja laatu sekä mitattavuus. Toisessa työpajassa (26.1.2017, 18 osallistujaa) työstettiin eteenpäin selvitystyössä muodostettuja alustavia suositusehdotuksia seuraavissa teemoissa: johtaminen ja toimintatavat, datainfrastruktuuri ja teknologia, juridiikka sekä ulkoinen toimintaympäristö ja osaaminen. Tavoitteena oli kattaa työpajoissa kaikki PESTL-näkökulmat.

4.2 Johtaminen ja ohjaus

Kappaleessa esitellään tietovarantojen hyödyntämisen kannalta merkitykselliset ohjauksen menetelmät: yleiset tavoitteet, tulosohjaus, tietotilinpäätös, datanavaussuunnitelma ja ai-neisto- tai datapolitiikka. Näitä hyödyntämällä voidaan päästä kohti yhtenäisiä, tehokkaita ja tuloksellisia toimintatapoja.

4.2.1 Valtioneuvoston digitalisaation tiedonhallinnan tavoitteet

Avoimen tiedon hyödyntämisen lähtökohtana tulee olla hyvä tiedonhallinta, jonka osana rakennetaan avoimen tiedon käytäntöjä. Tiedonhallinnalla tarkoitetaan toimenpiteitä, menetelmiä ja käytäntöjä, jotka mahdollistavat tietojen löytämisen, käsittelyn ja hyödyntämisen tietojen koko elinkaaren ajan. Tiedonhallinnan osa-alueita ovat mm. ydintietojen, metatietojen, asiati-tojen sekä avoimen datan hallinta ja ohjaus. Merkittävä osa avoimen datan kustannustehok-kaasta hyödyntämisestä perustuu tiedonhallinnan kokonaisuuden tehokkuuteen. Seuraavan sivun faktalaatikkoon on koottu keskeisiä avoimeen dataan liittyviä tavoitteita, linjauksia ja ohjeistuksia viime vuosilta.

Tietovarantoja on avattu avoimen tiedon ohjelman loppuraportin (2015) viitoittamalla tiellä, mutta esimerkiksi tieto-osaamisen ohjelmaa ei ole käynnistetty. Tietoa ja parhaita käytäntöjä on pyritty jakamaan hallituksen kärkihankkeiden puitteissa ja erilaisissa verkostoissa kuten Big Data Forum Finland⁶, MyData-allianssi⁷ ja uusi julkishallinnon data- ja analytiikkaosaajien verkosto. Avoimen datan portaalissa Avoindata.fi:ssä on julkaistu datan avaamisen ohjeet ja datan avaajille on järjestetty verkkokoulutusta. Avoin tiede ja tutkimus -hanke⁸ puolestaan on lisännyt tieto-osaamista tutkimuksessa ja muun muassa tutkimusaineistoja on avattu.

Helmikuussa 2016 julkistettiin digitalisoinnin periaatteet, jotka toimivat yhteisinä pelisääntöinä kaikkien julkisten palveluiden kehityksessä. Niissä linjataan muun muassa tiedon jakamista ja rajapintoja, toimintamalleja, nopeaa palvelukehitystä ja varautumista. Yksi yhdeksästä peri-

⁶ <http://www.tieke.fi/display/biff/Big+Data+Forum+Finland>

⁷ <http://www.mydata.fi>

⁸ <http://avointiede.fi/>

Pääministeri Juha Sipilän hallitusohjelmassa 2015–2019 todetaan: ”Luodaan avoimella datalla ja tietovarantojen paremmalla hyödyntämisellä edellytyksiä uusille liiketoimintatiedoille”. Hallitusohjelman yleisenä läpileikkaavana teemana on digitalisaatio, ja tiedon hyödyntäminen on esillä ohjelman muissakin kohdissa.

Valtioneuvoston periaatepäätös datan hyödyntämisestä liiketoiminnassa (2016) puolestaan kattaa datatalouden laajasti ja siinä on avoimen datan ja perustietovarantojen lisäksi tuotu vahvasti esille myös omadatan (MyData) ja massadatan (Big Data) merkitys. Periaatepäätöksen mukaan palvelu- ja sovelluskehityksen edistämiseksi tarvitaan yhteistyömalleja, kannustimia, avoimen tiedon kulttuuria ja kokeiluja. Datan keruu ja yhteiskäyttöisyys, yhtenäiset tietomallit ja -rakenteet, ohjelmointirajapinnat ja standardit sekä tiedon omistajuuden ja käyttöoikeuksien täsmennetty suhde tehostavat julkisen ja yksityisen sektorin yhteistoimintaa ja vauhdittavat kehitystä. Myös ihmiskeskeinen datan hallinta ja henkilötiedon käyttö nähdään mahdollisena kilpailukykytekijänä.

Julkisen hallinnon tietovarantojen avaamisen periaatteet -linjauksessa (2011) asetettiin tavoitteeksi, että ”Kaikki merkittävät julkishallinnon hallinnoimat julkiset tiedot ovat vuoteen 2020 mennessä saatavilla koneluettavassa muodossa, maksutta ja selkein käyttöehdoin ottaen huomioon tietosuoja, yksityisyyden suoja ja muut lainsäädäntöön perustuvat rajoitukset.” Perusteina olivat hyödyt kansantaloudelle, läpinäkyvyys sekä parantunut tehokkuus.

Avoimen tiedon ohjelma ehdotti loppuraportissaan (2015), että viranomaistiedon avaamista jatketaan, painotusta siirretään tiedon hyödyntämiseen ja kasvatetaan tieto-osaamista. Ohjelma ehdotti seuraavia toimenpiteitä, joilla voidaan tukea hallinnon tuottavuuden kasvua ja tiedon tehokasta hyödyntämistä yhteiskunnassa:

- 1. Jatketaan ja tuetaan tietovarantojen avaamista suunnitelmallisesti siten, että kaikki merkittävät julkiset tietovarannot ovat vuosikymmenen loppuun mennessä koko yhteiskunnan saatavilla koneluettavassa muodossa, maksutta ja selkein käyttöehdoin.*
- 2. Suunnitellaan keskeisten perustietovarantojen tuotannon, hallinnan ja palvelujen yleinen malli. Tavoitteena on perustietovarantojen maksuton ja tekninen saatavuus viranomaisille, sujuvat tietovirrat kuntien ja valtion välillä sekä tietojen hyödyntäminen laajasti yhteiskunnassa ottaen huomioon lainsäädännön rajoitukset.*
- 3. Kasvatetaan tieto-osaamista laatimalla poikkihallinnollinen kansallinen tieto-osaamisen ohjelma. Digitaalisessa tietoympäristössä toimimiseen ja selviytymiseen tarvitaan monipuolisesti tiedon tuotannon, jakamisen, avaamisen, analysoinnin ja hyödyntämisen taitoja.*

aatteesta onkin ”*avaamme tiedon ja rajapinnat yrityksille ja kansalaisille*”. Yhteinen tiedon hallinta -hanke⁹ (YTI) on osa hallituksen Digitalisoidaan julkiset palvelut -kärkihankekokonaisuutta. Hankkeen tavoitteena on luoda edellytykset julkisen hallinnon yhtenäiselle tiedonhallinnalle ja vahvistaa tiedon järjestelmällistä ja tavoitteellista määrittelyä ja käsittelyä. Hankkeen toimenpiteillä tehostetaan jo olemassa olevan tiedon monipuolista käyttöä ja digitalisoidaan julkisen hallinnon tiedon tuotannon taustaprosesseja.

Pääsääntöisesti haastateltavat tukivat valtioneuvoston peruslinjausta ”vuoteen 2020 mennessä tietovarannot avoimeksi”. Osa haastateltavista lisäsi kuitenkin, että tavoite on aikataulullisesti kunnianhimoinen ja joiltain osin haasteellinen toteuttaa. Lakisäätöiden tehtävien tuottaman datan avoimuus oli monelle ilmeinen tavoite. Haasteina tuotiin esiin varsinkin lainsäädännölliset tulkinnat, sirpaleinen ja osin vanha lainsäädäntö sekä henkilötietoa sisältävien tietoaineistojen hyödyntämiseen liittyvät haasteet. Uhkakuvina osa haastateltavista näki, että taloudelliset säästöt saattavat jatkossa haitata julkisen avoimen datan kehitystä siitä huolimatta, että datan

⁹ <http://vm.fi/hanke?selectedProjectId=22426>

avaamisella voi olla myös suoraan kustannuksia alentavia vaikutuksia. Myös datan avaamiseen liittyvän laadunhallinnan tärkeyttä ja osin haastavuutta korostettiin.

Joissakin organisaatioissa datan avaaminen on jo vakiintunut osaksi normaalia toimintaa, mutta haastattelujen perusteella monissa organisaatioissa datanavaussuunnitelmia ei ole päivitetty, ja lupaavan alun jälkeen tahti on hidastumassa. Keskeisten perustietovarantojen tuotannon, hallinnan ja palvelujen yleistä mallia ei ole otettu käyttöön, mutta sellaista ollaan valmistelemassa¹⁰. Tietovirtoja ei ole mallinnettu monissa organisaatioissa. Kuntien osuutta ei selvityksessä arvioitu.

4.2.2 Tulosohjaus ohjaamaan tietovarantojen avaamista ja hyödyntämistä

Tulossopimus on ministeriöiden ja virastojen välinen sopimus viraston toiminnallisista tavoitteista vuositason myönnettyjen määrärahojen puitteissa. Tulossopimukset tehdään tyypillisesti neljän vuoden jaksolle. Tulossopimuksille ei ole vakioitua yhteistä pohjaa tai sisältörakennetta, vaan ne ovat erilaisia eri hallinnonaloilla. Valtiovarainministeriö on määritellyt tulosohjauksen uudistamisen keskeisiksi tavoitteiksi, että ohjaus on tulevaisuudessa nykyistä strategisempaa, kevyempää, poikkihallinnollisempaa ja yhtenäisempää¹¹.

Avoim data oli usealla haastatellulla organisaatiolla kirjattuna tulossopimukseen, ja haastattelutavat pitivät tulosohjausta järkevänä keinona ohjata tietovarantojen avaamista ja hyödyntämistä sekä datapolitiikkaa ja tietopalveluita yleensäkin. Nykyisin tulossopimuksen sisältämät avoimen datan tavoitteet ovat usein tyyliltään toimenpiteitä kuvaavia, esimerkiksi ”avataan tietovaranto x”, eikä tavoitteita tai tavoiteltuja vaikutuksia kuvaavia.

Väestörekisterikeskuksen tulossopimuksessa todetaan:

”Väestötietojen hyödyntämistä lisätään avaamalla rakennustietoja. Rakennusten sijainti- ja osoitetiedoista muodostetaan julkinen ja maksuton aineisto yleisesti saataville. Myös ei-julkisten rakennus- ja huoneistotietojen maksuttomuutta edistetään.”

Patentti- ja rekisterihallituksen vastaavassa sopimuksessa puolestaan todetaan:

”Avoin data. Valtioneuvoston 3.3.2011 tekemän periaatepäätöksen mukaisesti maksujen perimisestä valtionhallinnon sisällä sekä kuntien ja valtionhallinnon välillä luovutaan. Lisäksi PRH avaa avoimen sähköisen rajapinnan, jonka käyttö olisi maksutonta. PRH voi kuitenkin edelleen periä tietopalvelumaksuja muussa toiminnastaan. PRH:lle aiheutuvat kustannukset ja tulomenetykset korvataan periaatepäätöksen mukaisesti. Avoimen datan palvelun käyttöönotto ja sen liittäminen kansalliseen palveluväylään vähentänee PRH:n tietopalvelujen maksullisten suoritteiden tuottoa suunnittelukaudella. Vaikutusta on vaikea arvioida luotettavasti etukäteen. Palvelun käyttöönoton vaikutuksia on jatkossa syytä seurata tulosneuvottelujen yhteydessä.”

Avoimen datan suunnitelmia tai toteutumistietoja voi löytyä esimerkiksi organisaation tulossopimuksesta, vuosikertomuksesta tai tietotilinpäätöksestä¹². Useiden organisaatioiden avoimen datan suunnitelmat puolestaan ovat muutaman vuoden takaa. Olisi perusteltua, että avoimen datan suunnitelmille olisi oma vakioitu osio sekä tulossopimuksessa että tietotilinpäätöksessä, sillä tietotilinpäätös dokumentoi, mitä on jo tapahtunut ja tulossopimus sisältää tavoitteet lähitulevaisuudelle.

¹⁰ <http://valtioneuvosto.fi/hallitusohjelman-toteutus/digitalisaatio/karkihanke1#toimenpide3>

¹¹ <http://vm.fi/hallintopolitiikka/ohjausjarjestelmat/tulosohjaus>

¹² https://www.trafi.fi/filebank/a/1463394812/9478dbd63f9d555bc3cd6f5b5bc99e54/20667-Trafin_julkaisu_14-2016_-_Tietotilinpaaotos_2015.pdf

4.2.3 Tietotilinpäätös tiedonhallinnan ja avoimen datan apuvälineenä

Tietotilinpäätös on organisaation oman tarkastelun tuloksena syntyvä raportti, joka antaa kokonaiskuvan organisaation tiedonhallinnan nykytilasta. Sen tavoitteena on laajentaa tilinpäätöksen kaltainen vuotuinen tarkastelu kattamaan myös organisaation tietovarannot, tietojohtamisen, tietojenkäsittelyn ja tietoturvallisuuden (Tietosuojavaltuutetun toimisto, 2012). Tietotilinpäätös ohjaa ajattelemaan tietoa resurssina ja jäsentämään tiedonhallintaa osana suurempaa kokonaisuutta. Se on suositteleva sekä sisällöltään että toimenpiteiltään. Vaikka esimerkiksi tietosuojavaltuutettu on suositellut tietotilinpäätöksen tekemistä jo vuosia, vain harva organisaatio on tehnyt sen. Selvityksessä mukana olleista organisaatioista tietotilinpäätöksiä olivat tehneet Väestörekisterikeskus (vuodesta 2010 alkaen), Verohallinto (vuonna 2013) ja Trafi (vuodesta 2015 alkaen).

Eri hallinnonalojen tietotilinpäätöksiä on vaikea vertailla keskenään, koska tietotilinpäätöksen sisältöä ei ole standardoitu ja niiden julkisuus vaihtelee. Tietotilinpäätöksen esikuvana toimiva taloudellinen kirjanpito on pitkälle standardoitu, millä saavutetaan vertailukelpoisuus. Kun julkishallinnon suorittama työ on erittäin suurelta osin tietotyötä, ei tietotilinpäätös lisää hallinnollista taakkaa, vaan vähentää sitä jopa merkittävästi pidemmällä aikavälillä. Tietotilinpäätöksen dokumentoidun tiedonhallinnan avulla voidaan lisätä virastojen, laitosten ja hallinnonalojen välistä yhteistyötä, vähentää päällekkäistä ja turhaa tekemistä sekä modernisoida toimintatapoja ja prosesseja.

Trafin vuoden 2015 tietotilinpäätöksessä (2016, s. 5) kuvataan:

”Trafin tietoa hyödynnettiin enemmän kuin koskaan, esimerkiksi tietoyksikköjä luovutettiin yrityksille ja muille viranomaisille noin 500 miljoonaa kappaletta, avointa dataa on ladattu jo yli 200 miljardia tietoriviä”. Lisäksi avointa dataa on käsitelty tavoitteena ja mietitty avoimen datan merkitystä laajemmin. Esimerkiksi todetaan, että ”viraston internet-sivustolle on avattu palautekanava, joka mahdollistaa avoimen datan ongelmatilanteiden, palvelukehityskysymysten ja tietoaaineistoon liittyvien kysymysten raportoinnin viraston käsiteltäväksi”.

4.2.4 Organisaatioiden aineisto- tai datapolitiikat

Aineisto- tai datapolitiikka on organisaation tietoaaineistojen hyödyntämistä edistävä johtamisen työväline. Datapolitiikka määrittelee painopisteet ja toimet, joilla organisaation tietovarantojen käyttöä edistetään. Vaikka kokonaisvaltaista ja yleistä datapolitiikkaa, jota avoimen tiedon ohjelman loppuraportissa peräänkuulutettiin, ei vielä olekaan, niin joissain organisaatioissa, kuten Terveyden ja hyvinvoinnin laitoksella ja Luonnonvarakeskuksella sellainen on määritelty¹³. Myös eri organisaatioiden toimialakohtaisilla yhteenliittymillä, kuten LYNETillä (luonnonvara- ja ympäristöalan tutkimusverkosto) ja Soterkolla (sosiaali- ja terveysalan asiantuntijalaitosten yhteenliittymä) on olemassa tai kehitteillä yhteinen datapolitiikka.

Seuraavan sivun laatikossa on Terveyden ja hyvinvoinnin laitoksen datapolitiikka, joka tekee selväksi organisaation tahtotilan ja suuntaviivat.

¹³ <https://www.luke.fi/palvelut/luken-aineistopolitiikka/>

THL:n datapolitiikka

THL:n datapolitiikka on tietoaineistojen tehokäyttöä edustava johtamisen ja käytännön työskente-
lyn väline. Datapolitiikka määrittelee painopisteet, joilla THL:n tietovarantojen käyttöä edistetään
vuosina 2016–2020.

Tietoaineistojen käyttöä tehostetaan yhteistyössä THL:n sisällä, kansallisesti ja kansainvälisesti.
Painotamme aineistonhallintasuunnitelman laatimista aineistojen elinkaaren alussa.
Kehitämme turvallista sähköistä tietoaineistojen käyttöä, säilyttämistä ja arkistointia.

Datapoliittiset toimet

Tietoaineistot näkyviksi

- Tietojen käytön lähtökohta on, että saatavilla on tietoa jo olemassa olevasta tiedosta, metatieto-
toa.
- Teemme aineistojen systemaattisesta kuvailusta ja metatietojen julkaisusta askelittain vakiintu-
neen toimintatavan.

Ajantasaisuutta, relevanssia ja asiakaskohtaista raportointia

- Arvioimme tietoaineistojemme ajantasaisuutta ja relevanssia rohkeasti.
- Hyödynnämme jo kerättyä tietoa emmekä tarpeettomasti kerää päällekkäistä tietoa.

Avoin data ja avoin julkaiseminen lähtökohdiksi

- Suosimme datan avaamista ja yhteistyötä kehittäjien kanssa.
- Julkaisemme THL:n keskeisten tietoaineistojen metatiedot, tilastotiedot ja perustulokset avoime-
na datana.

Tietoaineistot saataville

- Edistämme käyttörajoitteisten aineistojen saatavuutta laatimalla tietoaineistojen luovutusten ja
omistajuuden pelisäännöt.
- Tavoitteena on käyttö lupien ja tietoluovutusten kansallinen yhden luukun palvelu.
- Tiedonhallinnan koko elinkaari sähköiseksi.

(Lähde: <https://www.thl.fi/fi/tutkimus-ja-asiantuntijatyo/datapolitiikka>)

Taulukko 4.1 Kooste avoimen datan ohjauskeinoista

Organisaatio	Avoim data tai rajapinnat tulossopi- muksessa 2016	Avoimen datan suunnitelma laadittu	Tietotilinpäätös	Aineisto- tai datapolitiikka
Ilmatieteen laitos	Kyllä	2011–2012	Ei	Linjaus olemassa – ei verkossa
Kela	Ei	Ei (tekeillä 2017)	Ei	Ei
Liikennevirasto	Kyllä	2016	Ei	Ei
Liikenteen turvallisuusvirasto (Trafi)	Kyllä	2013, sekä osa LVM suunnitelmaa	2015 alkaen	Vaikeasti löydettävä
Luonnonvarakeskus	Kyllä	2016, liitteenä tulossopimuksessa	Ei	Verkkosivulla
Maanmittauslaitos (MML)	Kyllä	2012	Ei	Linjaus olemassa – ei verkossa
Patentti- ja rekisterihallitus (PRH)	Kyllä	Ei	Ei	Tekeillä
Suomen ympäristökeskus (SYKE)	Kyllä	2008 ohje tietoaineiston luovutta- misesta, 2016 lisenssi	Ei	Lynet
Terveystieteiden tutkimuskeskus (THL)	Ei	2016	Ei	Verkkosivulla
Tilastokeskus	Kyllä	2014	Ei	Vaikeasti löydettävä
Väestörekisterikeskus (VRK)	Kyllä	Ei	2010 alkaen	Ei
Verohallinto	Ei	Ei	2013	Ei

4.2.5 Kooste avoimen tiedon ohjaukeinoista

Taulukossa 4.1 on koostettu yhteen organisaatioiden tietovarantoihin liittyviä eri ohjaukeinoja. Suurimmassa osassa organisaatioita (9 kpl) tietovarantojen avaaminen oli tulossopimuksen osana. Vain kolme organisaatiota oli tehnyt tietotilinpäättöksen ainakin kerran. Päivitettyjä datan avaamissuunnitelmia oli alle puolella organisaatioista. Puuttuva tai useita vuosia sitten tehty avoimen datan suunnitelma ei välttämättä tarkoita, ettei dataa ole julkaistu, sillä datan avauksia voi olla kirjatuna esimerkiksi tulossopimukseen. Aineisto- tai datapolitiikka oli eksplisiittisesti kirjattuna vain kahdella organisaatiolla. Datan avaamissuunnitelma ja aineisto-/ datapolitiikka ovat potentiaalisesti näkyviä viestintäkeinoja sidosryhmille päin ja sikäli tärkeitä työkaluja tulossopimuksen ja tietotilinpäättöksen lisäksi.

4.3 Avoimen datan kustannukset ja lisäarvo

Tieto on organisaatioiden aineetonta pääomaa, joka liittyy organisaatioiden tehtävien hoitamiseen, johtamiseen ja toiminnan suorittamiseen (Moody & Walsh, 1999). Tietotyössä, jota eri hallinnonalojen virastot ja laitokset tekevät, tieto on toiminnan tärkein resurssi. Tieto tulee nähdä organisaation menestystekijänä ja yhä tärkeämpänä osana kokonaistoimintaa, eikä avointa dataa tulisi tarkastella vain erillisenä asiana. Kun toimitaan monimuotoisen tiedon kanssa kokonaisvaltaisesti, niin on hyvä ottaa huomioon erilaiset näkökulmat tietoon, joista yksi on avoimen datan näkökulma. Hyvällä ja tehokkaalla tiedonhallinnalla koko organisaation tieto on tehokkaasti ja turvallisesti käytössä. Kun tähän pyritään, tiedon arvo sekä organisaation että laajemmin koko yhteiskunnan pääomana nousee. Avoimen datan taloudellista näkökulmaa määrittää merkittävästi laitosten, virastojen ja ministeriöiden oma budjetointi ja hallinnonalan oma ohjaus. Nykyinen taloudellinen malli voi ohjata optimoimaan tulossopimuksessa käsiteltävää osakokonaisuutta yhteiskunnan kokonaisedun sijasta.

4.3.1 Datan avaamisen kustannukset

Datan avaamisen kustannuksiin vaikuttaa merkittävästi organisaation yleinen tiedonhallinnan tila ja kypsyys ajatella tietoa strategisena resurssina sekä käsitellä sitä koko elinkaaren ajan sähköisesti ja rakenteellisessa muodossa hallitusti. Arviot datan avaamisen kustannuksista vaihtelevat merkittävästi. Kustannusten laskemisessa on eroja tulkinnoissa ja sisällöissä, jonka takia niiden vertailu antaa ainoastaan suuntaa. Taulukossa 4.2 on jaoteltu kustannusten muodostumista datan avaamisprosessin vaiheiden mukaan. Taulukossa on lihavoidulla tekstilä kohdat, joissa syntyy merkittäviä kustannuseroja riippuen esimerkiksi tiedonhallinnan tilasta.

Datan avausprosessin ensimmäisessä vaiheessa esiintyi paljon eroja. Havainnollistava esimerkki tiedonhallinnan nykytilanteesta ja virastojen hallussa olevasta tiedon monimuotoisuudesta on Luonnonvarakeskus, jonka vanhimmat aineistot ovat 1800-luvun lopulta. Luonnonvarakeskus on perustettu vuonna 2015 neljän organisaation fuusiona. Sen tietoaineistot on kuvattu Radar-metatietopalveluun¹⁴, jossa (22.1.2017) oli 855 aineistokuvausta. Aineistoista osa on lakisääteisesti tuotettavia ja osa tutkimusaineistoja. Osa vanhimmista aineistoista on edelleen ”mustakantisissa vihkoissa” ja ”maastossa paperille kerättyä”, jolloin aineiston avaaminen voi olla jopa ”henkilövuosien työ”. Lisäksi ”tyypillisesti materiaalit pitää jalostaa tietotuotteeksi”, jotta niistä on avoimena datana hyötyä tutkimuskäyttöä laajemmin. Samankaltainen tilanne aineistojen suhteen on myös esimerkiksi Patentti- ja rekisterihallituksella. Se

¹⁴ <http://radar.luke.fi/>

aloitti toimintansa vuonna 1835 ja noin 180 vuoden aikana on tapahtunut paljon yrityskauppoja ja toimintojen myymisiä, ja lainsäädäntö on muuttunut kerättävän ja käsiteltävän tiedon osalta useita kertoja. Patentti- ja rekisterihallitus ei ole historiansa aikana digitalisoinut tai harmonisoinut tietoja tai prosesseja kokonaan.

Datan avausprosessin toisessa vaiheessa kustannuksia syntyy merkittävästi, jos materiaali ei ole digitaalisessa muodossa vaan esimerkiksi paperilla tai kuvina (skannatut paperit). Silloin se on ensin tallennettava rakenteellisessa muodossa, jonka kustannukset riippuvat materiaalin laajuudesta ja määrästä. Seuraavaksi aineistoa käsitellään, jolloin kustannuksissa on suuria eroja riippuen aineiston tyypistä ja tarpeista. Henkilötietojen poistaminen voi olla hankalaa, jos ne voidaan tulkita eri tavoin. Tässä vaiheessa tulee huomioida metatietojen lisääminen (tietomalli ja aineiston saatavuudesta kertominen) sekä ohjelmointirajapinnan, palvelulupauksen ja hallinnointimalliin kuvaaminen. Tämä vaikuttaa kustannustehokkuuteen. Kun metatiedot ovat riittävän hyvät, datan hyödyntäjän ei tarvitse arvailla tietomallia eikä ihmetellä, miten data-aineistoa pitäisi tulkita.

Datan avausprosessin kolmannessa vaiheessa kustannustehokkuutta luo se, että hyödynnetään jo olemassa olevia palveluita datan julkaisemisessa, eikä välttämättä luoda omaa uutta palvelua sitä varten. Toisaalta haastatteluissa oman palvelun luomista perusteltiin oman osaamisen ja toiminnan markkinoinnilla. Tiedon tuotantoprosessit tiedon keruusta sen avaamiseen voivat toimia kustannustehokkaimmin lähellä tiedon varsinaista tuottajaa. Yhtenäinen palvelu tai jakelualusta voi muodostaa loogisen, helposti hyödynnettävän kokonaisuuden tietyn sisältöiselle ja tyyppiselle aineistolle. Palvelun toteuttamiselle tietyllä tavalla voi olla myös lainsäädännöllisiä perusteita.

Taulukko 4.2 Julkisen avoimen datan kustannustyyppijä avausprosessin työvaiheittain ja pääpiirteittäin (Avoimen datan opasta^a soveltaen)

Vaihe	Tehtävät	Kustannuksia
0 Tietoisuus avoimesta datasta	Tyypillisesti julkilausuma hallinnon läpinäkyvyydestä riittävän vakuuttavan tahon toimesta	Sitoutuminen esimerkiksi Open Government Partnership -ohjelmaan ^b
1 Avattavan aineiston valitseminen	1a. Organisaation tietoaineistojen kartoittaminen 1b. Luokittelu (avattavissa oleviin, lisäselvitystä vaativiin ja salassa pidettäviin) 1c. Luettelon julkaisu kaikista organisaation tietojärjestelmistä ja olemassa olevista tietovarannoista	Oma asiantuntijatyö Oma tietopalvelu- tai ICT-työ Oma juridinen työ Tiedonhallintaan liittyvä työ
2 Datan valmistelu avaamista varten	(2a. Siirtäminen digitaaliseen muotoon) 2b. Aineiston käsitteleminen (yhdistäminen, muokkaaminen, siivoaminen ja jalostaminen) 2c. Henkilötietojen ja muiden ei-julkisen tietojen poistaminen 2d. Tallentaminen julkaistavaa tiedostomuotoon 2e. Metatietojen lisääminen 2f. Ohjelmointirajapinnan dokumentointi 2g. Käyttöluvan valitseminen (esimerkiksi CC BY 4.0 tai CC0)	Oma asiantuntijatyö Oma tietopalvelu- tai ICT-työ Ulkoa ostettu ICT-työ
3 Datan julkaisu	3a. Omassa verkkopalvelussa (tai avoindata.fi) 3b. Linkittäminen muihin verkkopalveluihin (datakatalogit)	Oma asiantuntijatyö Oma tietopalvelu- tai ICT-työ Oma juridinen työ
4 Tiedottaminen ja vuorovaikutus	4a. Viestintä datan avaamisesta 4b. Vuorovaikutus datan hyödyntäjien kanssa (kilpailut, datan hyödyntäjille suunnatut tapahtumat, hackathonit, jatkuva datan hyödyntäjien tuki ja kumppanikoodarit)	Oma asiantuntijatyö Oma tietopalvelu- tai ICT-työ Markkinointi ja viestintä
5 Dokumentointi, jatkokehittäminen ja ylläpito (ks. myös seuraava alaluku)	5a. Avausprosessin dokumentointi ja hyötyjen reflektointi 5b. Raportointi sidosryhmille ja uusille työntekijöille 5c. Ylläpitokustannukset 5d. Tukipalvelut 5e. Jatkokehittäminen	Oma asiantuntijatyö Oma tietopalvelu- tai ICT-työ Viestintätyö Ulkoa ostettu ICT-työ Ulkoa ostettu vuorovaikutustyö Raportointityö

^a <https://www.avoindata.fi/fi/opas>

^b <https://www.opengovpartnership.org/>

Datan avausprosessin neljännessä vaiheessa kustannustehokkuuseroja tulee siitä, miten hyvin ja laajasti viestitään ja miten datan hyödyntäjiä kannustetaan hyödyntämään avattua dataa. Esimerkiksi kilpailujen, tapahtumien ja hackathonien järjestäminen maksaa ”ylimääräistä”, mutta todennäköisesti edistää avattua dataa hyödyntävien sovellusten kehittämistä.

Datan avausprosessin vaiheesta viisi on vielä suhteellisen vähän kokemusta, koska dataa on avattu vasta viime vuosina. Ylläpitokustannukset muuttuvat teknologisen kehityksen ja käytön muutoksen myötä. Tukipalveluissa on eroja henkilöresurssoinnissa. Jatkokehittämisen kustannuksia pystytään arvioimaan vasta myöhemmin (esimerkiksi 2020-luvulla) luotettavammin.

Haastattelujen perusteella avoimen datan julkaisuprosessi voi olla ensimmäisellä kerralla työläs (erään haastateltavan mukaan jopa miljoonia euroja), riippuen organisaation tiedonhallinnan tilanteesta. Tietokannoissa olevaa raakadataa jalostetaan ja julkaistaan tietotuotteina, esimerkiksi avoimena datana tai tietopalveluissa. Tyypillisesti tiedon avaamisessa selvitetään erilaisia lainsäädännöllisiä asioita, tehdään muutoksia tietojärjestelmiin ja rakennetaan tiedon avaamisen prosessi. Kun prosessi on kerran rakennettu, sitä on helpompi ja nopeampi toistaa. Prosessia voidaan kuitenkin joutua tarkastelemaan uudelleen esimerkiksi lainsäädäntö-, ohjaus-, resurssi- ja säännösmuutosten seurauksena. Haastattelujen perusteella datan avaaminen ei ensimmäisen kerran jälkeen vaadi merkittävästi resursseja, vaan jo avoimen tietoaineiston päivittäminen (esimerkiksi kuukausittain tai vuosittain) voi olla hyvinkin pieni, esimerkiksi alle päivän tai muutamien päivien työ. Haastateltavien organisaatioiden tiedonhallinta ja tietovarannot olivat kuitenkin keskenään erilaisia, joten resursointi ja työmäärätkin vaihtelevat tietoaineistokohtaisesti.

Kun tieto käsitellään koko elinkaaren ajan sähköisessä ja rakenteellisessa muodossa, on avoimen datan kustannustehokas hyödyntäminen optimaalista. Eräs haastateltava nosti esiin mahdollisena kehityskohteena talous- ja tilinpäätöstietojen sähköisen käsittelyn esimerkiksi XBRL-muodossa (eXtensible Business Reporting Language), jonka käyttö on Tanskassa pakollista. Suomessa yritykset puolestaan tyypillisesti lähettävät yhtiöjärjestykset edelleen paperilla PRH:lle, joka skannaa ne kuvana tietojärjestelmään. Yritystietojen käsittelemisessä koko elinkaaren ajan sähköisesti ja avaamisessa rakenteellisessa muodossa on paljon hyödyntämättömiä mahdollisuuksia. Yrityksiä koskevat tiedot ovat selkeästi korkean prioriteetin tietoja (Sasse et al., 2017, s. 21). Tuki XBRL-tietoformaatile on myös mainittu Tilastokeskuksen tulostavoitteissa.

4.3.2 Taloudellinen resursointi datan avaamiseen ja ylläpitoon

Datan avaaminen on selkeä 2010-luvun megatrendi. Tietoa kannattaa pääsääntöisesti jakaa mahdollisimman tehokkaasti laajalle, jotta siitä saatava hyöty maksimoituu ja se hyödyttää mahdollisimman monia. Avoimen datan kustannuksia tulee tarkastella kokonaisvaltaisesti koko tiedon elinkaaren ajan. Elinkaarikustannuksia voivat aiheuttaa muun muassa:

- avoimen datan palveluiden kuormitus
- luvattu palvelutaso (SLA) jakelussa
- ylläpito- ja tukipalvelut
- mahdolliset tietoturvaongelmat
- muuttuvat formaatit ja standardit
- muuttuvat velvoitteet ja lainsäädäntö
- tiedon korjaaminen, täydentäminen ja päivittäminen

Yhä suurempi osa avoimesta datasta julkaistaan ohjelmointirajapintojen kautta. Niiden rakentaminen on tyypillisesti kalliimpaa kuin tiedostopohjainen jakelu. Ohjelmointirajapintojen hyötyjä ovat reaaliaikaiset datavirrat, säästyneet tietoliikennekustannukset, tarkempi datan käytön seuranta, datan julkaiseminen useilla eri metadatoilla ja useissa formaateissa sekä data-aineistojen suurempi räätälöintimahdollisuus. Yhä useampi esine sisältää sulautetun ohjelmiston, kytkeytyy internetiin ja hyödyntää ohjelmointirajapintoja. Avointa dataa tarjoavaa rajapintaa voivat modernissa kotitaloudessa hyödyntää esimerkiksi älypuhelimet, tabletit, pelikonsolit, autot, jääkaapit, pesukoneet, valaistusjärjestelmä, lukot ja monet muut laitteet ja niiden ohjelmistot (Greengard, 2015). Siten rajapintoihin saattaa kohdistua valtava määrä kyselyitä, ja niiden resurssitarpeet voivat kasvaa huomattavasti varsinkin, jos ko. laitteet käyttävät rajapintapalveluita tuhmailevasti. Tietoturvanäkökulmasta tarkasteltuna avointa dataa jakavia ohjelmointirajapintoja vastaan voidaan tehdä esimerkiksi hajautettuja palvelunestohyökkäyksiä.

Tyypillisesti dataa avataan ja siihen liittyviä elinkaarikustannuksia hoidetaan jo olemassa olevien resurssien avulla. Toisinaan voidaan myöntää joko kertaluontoista avustusta datan avaamiseen tai budjettiavustusta tulojen menetykseen. Tosin eräs haastateltava totesi, että ”projektirahoitusta voi hakea ja saada palveluiden kehittämiseen, mutta ei datan avaamiseen”. Monen haastatellun organisaation budjettirahoitus oli vähentynyt ja yleisiä resursseja on käytössä aiempaa vähemmän. Useimmilla organisaatioilla on hankkeita ja velvoitteita alkamassa tai meneillään liittyen mm. kansalliseen palveluväylään¹⁵ sekä EU:n tietosuojasetukseen. Vaikka nämä ovatkin periaatteessa pääosin erillään avoimesta datasta, käytännössä osin samat henkilöresurssit ovat kiinnitettyinä kaikkiin tiedonhallinnan eri osa-alueiden hankkeisiin, joten on mahdollista, että resurssointi tietovarantojen avaamiseen heikentyy. Haastatelussa mainittiin myös, että datan avaamisen on pelätty johtavan lisääntyvään työmäärään, esimerkiksi lisääntyneen tuen kautta samalla, kun avoimen datan käyttö kasvaa.

4.3.3 Datan avaamisen taloudelliset vaikutukset

Kun Yhdysvaltain presidentti Ronald Reagan määräsi satelliittipaikannusdatan avattavaksi vuonna 1983, ei kukaan osannut ennustaa, että 30 vuotta myöhemmin tätä dataa käytettäisiin niin monin eri tavoin, että arviolta 3,3 miljoonaa työpaikkaa Yhdysvalloissa perustuu datan hyödyntämiselle (Pham, 2011; Economist, 2013). Tietoaineiston avaaminen tehostaa työajan käyttöä, kun esimerkiksi laskutuksesta ja erilaisista sopimusten luomisesta ja hallitsemisesta luovutaan. Laskutuskustannukset pienenevät ja yhteistyö sujuu entistä helpommin. Laskutukseen liittyvien tietojärjestelmien käytöstä voidaan luopua, mutta toisaalta avoimen datan jakelun tietojärjestelmäkustannuksia voi tulla lisää datan käytön kasvaessa. Tietotuotteiden tulot ja niihin tarvittava räätälöintityö vähentyvät. Myös sopimuskulut laskevat.

Ulkoisina vaikutuksina päällekkäinen työ vähenee, datan käyttö lisääntyy, yhteistyömahdollisuudet ja tehokkuus paranevat, sekä syntyy uusia yrityksiä ja liiketoimintaa. Toisaalta datan avaus muuttaa liiketoimintaympäristöä, jolloin liiketoimintaa voidaan myös menettää. Taulukossa 4.3 on esitetty arvio avoimen datan tuottamista hyödyistä EU:n alueella vuosina 2016–2020.

Valtiovarainministeriön avoimen datan pilottiorganisaationa toiminut Suomen Ympäristökeskus (SYKE) kertoi haastattelussa, että vuonna 2008 viraston paikkatietoaineistopaketeista maksettiin 300 kertaa ennen datan avaamista. Avoimen datan julkaisun jälkeen käyttö alkoi

¹⁵ Vuoden 2017 heinäkuun alkuun mennessä viranomaisten on tuotettava keskeisten palveluiden julkiset kuvaustiedot kansalliseen palvelutietovarantoon siitä edelleen palvelunäkymissä hyödynnettäväksi. Viranomaisten on tarjottava rekistereihinsä merkittyjä luonnollista henkilöä tai organisaatiota koskevia tietoja palvelunäkymässä ja 1.7.2017 otettava käyttöön viestinvälityspalvelu. 2018 alkuun mennessä otetaan käyttöön tunnistuspalvelu ja verkkomaksamisen kokoamis- ja hallinnointipalvelu.

Taulukko 4.3 Arvio avoimen datan taloudellisesta merkityksestä Euroopan unionin alueella vuosina 2016–2020

Markkinoiden koko	Työpaikat	Kustannussäästöt julkiselle sektorille	Tuottavuuden parantuminen
Kumulatiivinen markkinoiden koko on 325 miljardia euroa	100 000 työpaikkaa avoimessa datassa vuoteen 2020 mennessä	1,7 miljardia euroa (28 EU:n jäsenmaalle)	7 000 ihmishenkeä säästyy nopeamman reagoinnin seurauksena (ensiapu)
Avoimen datan markkina kasvaa 36,9 % yhteensä 75,7 miljardiin euroon vuoteen 2020 mennessä	Avoimen datan työpaikkojen määrä kasvaa 25 000:lla		2 549 tuntia säästyy parkkipaikan etsimisestä

Lähde: Berends et al., 2017, s. 7.

kasvaa merkittävästi. Jo vuonna 2008 aineistolatauksia oli noin 3 000 ilman suurta markkinointipanostusta. Vuoteen 2015 mennessä kasvua oli noin 20 000 käyttäjään ja 47 000 lataukseen. Rajapintapalveluiden käyttö 2012–2015 kasvoi puolestaan lähes nelinkertaiseksi (n. 23 000 → 93 000). SYKE:n kokemuksen mukaan ”pienikin maksu (datan hyödyntäjille) vähentää käyttöä”.

Joillekin julkisen sektorin tiedon tuottajille tai rekistereiden ylläpitäjille datan avaamisesta on tullut tulonmenetyksiä, kun datan ja tietotuotteiden myynnistä on luovuttu tai sitä on vähennetty. Vastaavasti räätälöityjen palveluiden myynti on vähentynyt. Samalla datan avaaminen on vähentänyt laskutukseen ja käyttö sopimukseen liittyvää työtä ja tehostanut henkilöstön työajan käyttöä sekä datan tuottajan että hyödyntäjän organisaatioissa. Virastojen välinen laskutus ja sopimusten tekeminen datan jakamisesta on jäänyt pois.

Eräs haastateltava arvioi, että summaltaan pienten laskujen lähettäminen, perintä ja koko laskutustoiminnan pyörittäminen tekee tiedonluovutustoiminnasta organisaation sisällä jopa tappiollista. Summaltaan pienten laskujen maksaminen unohtuu asiakkailta helposti ja niiden siirtäminen perintään kuluttaa merkittävästi henkilöstöresursseja. Haastateltavan arviolaskelman perusteella olisi taloudellisesti järkevää luopua laskutustoiminnasta ja jakaa data avoimesti muun yhteiskunnan käyttöön. Laskutuksesta syntyy myös kustannuksia henkilöstö-, postitus- ja tietojärjestelmäkuluina.

Maanmittauslaitoksen paikkatietoaineiston avaaminen on parantanut julkisen sektorin yhteistoimintaa mm. vähentämällä lupamenettelyä ja julkisen sektorin keskinäistä laskutusta. Osa maksuista, joka näkyi ennen Maanmittauslaitoksen tulona, näkyi toisissa virastoissa menoina. Toisaalta datan avaamisen kokonaiskustannukset olivat ennakkosuunnittelua suuremmat kuin saadut kompensatiot, koska avaamisen seurauksena esimerkiksi räätälöityjen palveluiden myynti on vähentynyt. Tarkkoihin taloudellisiin lukuihin liittyy suurta tulkinnallisuutta tällä osa-alueella, joten niitä voidaan korkeintaan pitää suuntaa-antavina.

Osa haastateltavista oli sitä mieltä, että vaikka avoin data on vapaasti saatavilla, se voi olla luonteeltaan sellaista, ettei sitä pysty hyödyntämään ilman erikoisosaamista. Tällöin dataa avaava organisaatio voi myydä palveluna myös datan hyödyntämisen osaamista datan tulkintaan ja ymmärtämiseen. Toisena vaihtoehtona mainittiin mahdollisuus myydä palveluna räätälöityjä avoimen datan aineistoja.

Väestörekisterikeskus ja Patentti- ja rekisterihallitus ovat esimerkkejä nettobudjetoiduista virastoista ja perusrekistereiden ylläpitäjistä. Ne saavat tuloja tietotuotteista ja -palveluista, jotka perustuvat niiden tietovarantoihin. Esimerkiksi PRH saa rekistereissä olevasta tietojen myynnistä noin 2,5–3 miljoonaa euroa ja Väestörekisterikeskus noin 10 miljoonaa euroa vuodessa.

Nettobudjetoitujen virastojen tietovarantojen laajempi avaaminen luo taloudellisessa mielessä sekä mahdollisuuksia että uhkia. Näitä kannattaisi selvittää ja mallintaa tarkemmin mm. skenaarioanalyysin ja standardoitujen laskentatapojen avulla.

Maanmittauslaitoksen datan avaaminen on luonut uutta liiketoimintaa. Tosin on myös yrityksiä, kuten esimerkiksi Karttakeskus, jotka menettivät huomattavan osan liiketoiminnastaan datan avaamisen seurauksena. Trafi puolestaan toimii aktiivisesti yhteistyössä kasvuyritysten kanssa. Trafin avaamaa dataa käyttää yli 70 yritystä. On useita yrityksiä, jotka jalostavat osin tai kokonaan käyttörajoitteista dataa ja myyvät palveluita, toisinaan myös julkishallinnolle itselleen. Esimerkiksi yritysrekisterin laajemmalla avaamisella voi olla julkisen sektorin dataa rikastaville yrityksille sekä liiketoiminnallista haittaa että hyötyä. Yleisesti tietovarantojen lupa-käytäntöjen muuttaminen ja datan avaaminen luo kuitenkin uutta kilpailua.

Henkilötiedot ovat yksi merkittävimmistä tulevaisuuden liiketoimintaa muokkaavista voimista (World Economic Forum, 2013) ja niihin liittyvien palvelujen kokonaismarkkinan on arvioitu kasvavan Euroopassa jopa 1 000 miljardiin euroon vuonna 2020 (Boston Consulting Group, 2012). Vastaavasti avoimen datan potentiaaliseksi suoraksi hyödyksi on arvioitu 40 miljardia ja välillisiksi 100 miljardia euroa vuosittain EU-alueella (Vickery, 2011). Arviot ovat suuntaa antavia, mutta kokoluokaltaan henkilötiedon taloudellinen potentiaali on näin arvioituna jopa kymmenkertainen avoimen datan potentiaaliin verrattuna. Siinä mielessä on tärkeää miettiä avoimen datan hyödyntämismahdollisuuksia yhdessä henkilötiedon kanssa, kuten monet haastateltavat totesivat.

4.4 Tehokasta datan hyödyntämistä tukevat toimintatavat

Luku tiivistää löydökset toimintatavoista, jotka edistävät datan tehokasta hyödyntämistä. Erityisesti mahdollisuuksia nähtiin asiakaslähtöisessä, kokeiluhenkisessä ja järjestelmällisessä vuorovaikutuksessa datan tuottajien ja hyödyntäjien kesken. Vuorovaikutus hyödyntäjien kanssa on olennaista myös datan avaamisen priorisoinnin ja osaamisen kehittämisen kannalta. Avoimilla toimintatavoilla voidaan päästä irti myös toimittajalukoista.

4.4.1 Datan avaamisen priorisointi kysyntä- ja vaikuttavuuslähtöisesti

Euroopan unioni on määritellyt viideksi tärkeimmäksi avoimen datan osa-alueeksi tilastot, paikka-, liikenne- ja yritystiedot sekä maapalloa koskevan havaintotiedon (Sasse et al., 2017, s. 21). Data-aineistojen avaamisen priorisoinnin suhteen euroopanlaajuinen, julkisen datan hyödyntämistä edistävä EU-hanke SharePSI¹⁶ nostaa esiin erityisesti ns. korkean lisäarvon data-aineistot ja niiden tunnistamiseen liittyvät toimet.

Datan hyödyntäjän näkökulmasta data-aineiston arvo perustuu erityisesti käyttö- ja uudelleenkäyttöpotentiaaliin, joka voi johtaa uusiin liiketoimintamalleihin. Tiedon julkaisijan näkökulmasta korkean lisäarvon data- tai tietoaaineistoja ovat puolestaan aineistot:

- jotka lisäävät läpinäkyvyyttä
- joiden julkaisuun on laillisia velvoitteita
- jotka liittyvät suoraan organisaation julkiseen tehtävään
- joiden julkaisusta avoimena datana on suoria kustannushyötyjä
- joilla on laaja käyttäjäkunta. (Archer et al., 2014).

¹⁶ <https://www.w3.org/2013/share-psi/>

Archer et al. (2014) ovat selvityksessään käyneet läpi EU-instituutioiden eri data-aineistoja ja arvioineet niiden potentiaalista lisäarvoa ym. kriteerien perusteella. Joidenkin tutkimusten mukaan esimerkiksi julkisuuslain mukaiset tietopyynnöt antavat viitteitä, mitä dataa kannattaisi avata (Reinvent Albany, 2014). Myös muu epäformaalimpi palaute tiedon hyödyntäjiltä voi olla relevanttia. Myös valtiovarainministeriö on ohjeistanut priorisointia datan avaussuunnitelmien laadinnassa.

Datan avaamista periaatepäätöksen jälkeen on johdettu vaihtelevasti. Organisaatioilla on ollut datan avaamiseksi omia suunnitelmiaan, joita on laadittu ja päivitetty vaihtelevasti, osin hyvinkin tarkkaan. Monella organisaatiolla ei ole varsinaista avoimen datan suunnitelmaa, mutta tiedon avaamisen kehittäminen on kirjattu tulostavoitteisiin.

Osin tietoaineistojen avaamista on priorisoitu avaamisen helppouden perusteella. Esimerkiksi Väestörekisterikeskus on avannut nimidataa eli kuinka yleisiä tietyt henkilöiden nimet ovat väestön joukossa. Tämä on helppo ja kysytty aineisto, mutta voidaan kysyä, kuinka merkityksellinen se on, eli onko sen vaikuttavuus suurta. Pelkkä kysyntä ja asiakaslähtöisyys eivät siis itsessään välttämättä lisää vaikuttavuutta. PRH on avannut YTJ-tietoja ja kaupparekisterin kuulutustietoja. Monella toimijalla on lakisääteisten tehtävien tuottamaa dataa erittäin monipuolisesti ja laajasti avattu ja meneillään on esimerkiksi tutkimustiedon avaaminen (esim. SYKE, Luonnonvarakeskus ja Ilmatieteen laitos). Luonnonvarakeskus pyrkii avaamaan dataa, jota kysytään paljon, jotta vähennetään omaa työtaakkaa dataa avaamalla. Kun uusia tietojärjestelmiä hankitaan, avoin data tai avoimet rajapinnat on usein määritelty osaksi toteutusta eikä avaamista siten välttämättä tarvitse erikseen priorisoida (esim. Liikennevirasto).

Data-aineistojen avaamisen priorisointi työ- ja elinkeinoministeriössä

Työ- ja elinkeinoministeriön hallinnonalalla ja ministeriössä avataan:

1. Ensi tilassa teknisesti helposti avattavissa olevat tietovarannot, joiden avaamiselle ei ole esteitä eikä merkittäviä kustannus- tai toimenpidevaatimuksia. Tarkistetaan käyttöehdot.
2. Muutetaan maksuttomiksi sellaisia tietovarantoja, joiden muuttaminen ei tuota merkittäviä budjettivaikutuksia.
3. Parannetaan jo avattujen tietovarantojen käytettävyyttä rakentamalla rajapinnat koneluettavuudelle.
4. Tunnistetaan tietovarannoista ne, joiden avaaminen liittyy viraston tietojärjestelmämuutokseen/-uusimiseen ja aikataulutetaan niiden avaaminen tarkemmin.
5. Tunnistetaan tietovarannot, joiden jakeluun halutaan käyttää valtion yhteistä alustaa, väylää tai järjestelmää.
6. Kootaan tarkemmat tiedot tarvittavista avaamiseen ja maksuttomuuteen liittyvistä säädösmuutoksista.
7. Selvitetään erityisesti PRH:n ja GTK:n osalta, mitä tietovarantoja avataan, millä käyttöehdoilla ja missä aikataulussa. ELY-keskusten tietovarantokartoitus ja tietoaineistojen avaamisen suunnittelu tulisi toteuttaa osana eELY-ohjelmaa yhteistyössä ELY-keskuksia ohjaavien tahojen kanssa. Muiden virastojen osalta jatketaan tietovarantojen avaamista niin, että ilmoitettuja aikatauluja pyritään aikaistamaan.

4.4.2 Datan avaajien ja hyödyntäjien välinen vuorovaikutus

Säännöllinen ja suora vuorovaikutus tiedon tuottajien ja tiedon hyödyntäjien välillä on määriteltä yhdeksi parhaaksi käytännöksi mm. SharePSI-verkoston Best Practices¹⁷ -oppaassa, ja se on myös yksi olennainen keino priorisoida tietoaineistojen avaamista. Lisäksi SharePSI suosittelee mm. sähköpostilistoja, virtuaaliyhteisöjä ja verkkopohjaisen yhteistyön mahdollistavien työkalujen käyttöä, jotka tukisivat vahvaa vuorovaikutusta ja mm. liiketoimintamahdollisuuksien ristipölytystä ja yhteiskehittelyä.

Suomessa yksi esimerkki avoimen datan verkkoyhteisöstä on Facebookissa toimiva Finnish Open Data Ecosystem¹⁸ -ryhmä, jossa on yli 4 000 jäsentä – monipuolisesti mm. julkishallintoa, kansalaisaktivisteja ja yritysten edustajia. Voidaankin pohtia, voitaisiinko tällaisen yhteisön asema formalisoida ja miten vuorovaikutusta kyseisen yhteisön kanssa voisi tuloksellisesti lisätä? Voisiko yhteisön koordinointiin ja fasilitointiin ottaa strategisemmän otteen ja tulisiko kehittäjä- ja hyödyntäjäyhteisön olla julkishallinnon vai hyödyntäjien koordinoima?

Myös säännöllisiä kohtaamisia datan avaajien ja hyödyntäjien välillä tarvitaan. Tästä hyvä esimerkki on Helsinki Region Infosharen avokonttorikäytäntö (Helsinki Loves Developers¹⁹), jossa datan avaajat ja hyödyntäjät kohtaavat toisensa kerran kuussa teemoittain vaihtelevassa kolmen tunnin työpajassa. Helsinki Region Infoshare ja 6Aika-hanke myös ilmoittavat mahdollisille datan hyödyntäjille etukäteen tulevista suunnitelmista datan avausten suhteen²⁰.

Haastateltujen piirissä tuntemus datan hyödyntäjistä ja vuorovaikutus kehittäjien kanssa oli vaihtelevaa. Tärkeät käyttäjäryhmät tunnetaan yleensä, samoin yksittäiset dataa hyödyntävät sovellukset tai yritykset, mutta tarkkaa tietoa käyttäjistä ei ole. Moni organisaatio on ollut mukana esimerkiksi sovelluskilpailuissa (mm. Apps4Finland ja Open Finland Challenge) tai hackathon-tapahtumissa (mm. Ultrahack, IndustryHack, Junction ja omat hackathonit). Liikennevirasto on puolestaan käyttänyt Google Groups -ryhmiä keinona kerätä datansa hyödyntäjiä yhteen. Usean haastateltavan organisaation avoimen datan asiantuntijoita osallistuu mainitun Finnish Open Data Ecosystem -ryhmän keskusteluihin aktiivisesti.

Työpajoissa yhtenä tärkeimmistä avoimen datan haasteista nähtiin asiakaslähtöisyyden puute. Sen sijaan, että ”vain avataan dataa”, pitäisi datan avaamisen, julkaisemisen ja muun toiminnan olla enemmän asiakaslähtöistä. Vuorovaikutuksen lisäämiseen datan avaajien ja hyödyntäjien välillä esitettiin ratkaisuksi yhteisten tapahtumien, kuten esimerkiksi hyvin rajattujen hackathonien, kilpailujen ja muiden tapahtumien järjestämistä. Tapahtumat ovat myös keino kehittää datan avaajien omaa osaamista sen lisäksi, että ne sujuvoittavat tiedon hyödyntämistä. Haastatelluista ainakin Verohallinto, Kela, Tilastokeskus, THL, PRH, Liikennevirasto ja Trafi ovat kannustaneet datan hyödyntäjiä käyttämään niiden avaamaa dataa ja rajapintoja hackathoneissa. Niiden kokemukset tapahtumista olivat hyviä, mutta ne eivät ole vielä vakiintuneet säännölliseksi toiminnaksi.

Myös ketterä ja käyttäjälähtöinen kokeilukulttuuri nousi työpajoissa vahvasti esiin tavoiteltavana toimintatapana. Tähän sisältyvät liiketoiminnan sekä avaamistapojen ja -mallien pilotointi. Työpajassa todettiin, että kokeilukulttuuri vaatii johdon sitoutumisen sekä iteraatiivisen ja käyttäjälähtöisen toimintamallin. Kokeilukulttuurilla viitattiin mm. siihen, että olisi hyvä löytää keinoja vaikkapa hackathonien jälkeen tehdä toimeksiantoja kehittäjille. Tässä

¹⁷ <https://www.w3.org/2013/share-psi/bp/>

¹⁸ <https://www.facebook.com/groups/fi.okfn/>

¹⁹ <http://www.hri.fi/fi/hri-projekti/hel-loves-dev-avoin-konttori/>

²⁰ <http://www.hri.fi/fi/hri-projekti/tyon-alla/>, <http://www.databusiness.fi/fi/avoin-data/roadmap/>

voitaisiin hyödyntää innovatiivisia hankintoja, johon uusi hankintalaki antaa aiempaa paremmat mahdollisuudet.

4.4.3 Avoimuudella irti toimittajalukoista

Avoim data, avoimet ohjelmistorajapinnat, ohjelmistojen avoin lähdekoodi ja avoimet lisenssit voivat vapauttaa julkisen sektorin ns. toimittajalukoista, eli vähentää toimittajariippuvuutta. Samalla ne tuovat julkiselle sektorille vastuun hallinnoida datan avaamista ja käyttöä sekä koordinaida datan käyttöä tukevien ohjelmistojen kehittämistä ja ylläpitoa (Parviainen et al., 2017).

Organisaatioissa käytetyillä ohjelmistoilla on vaikutusta siihen, kuinka tehokas avoimen datan julkaisuprosessi on. Tähän liittyvää ohjelmisto-osaamista ei kannata ulkoistaa kokonaan, vaan ohjelmistotyön ostamisen tueksi tarvitaan riittävää ymmärrystä ohjelmistojen toiminnasta.

Toimittajalukkojen välttämiseksi avointa lähdekoodia ja ohjelmointirajapintoja tulisi hyödyntää julkisessa hallinnossa. Avoimesta lähdekoodista on olemassa oma JHS-suositus (JHS 169 Avoimen lähdekoodin ohjelmien käyttö julkisessa hallinnossa²¹). Liikennevirasto esimerkiksi tilaa kaikki räätälöidyt ohjelmistot lähtökohtaisesti avoimena lähdekoodina. Avoin lähdekoodi voi toimia yhdessä avoimen datan kanssa esimerkiksi kuten Ilmatieteen laitoksella: avoimen datan palvelin ja dataa käsittelevät ohjelmistot on julkaistu avoimella lähdekoodilla²². Näin kehittäjien on helpompaa (hyödyntämällä avointa koodia) ottaa data käyttöön, ja toisaalta itse datan prosessointi voidaan hajauttaa. Avoin lähdekoodi siis tukee avoimen datan käyttöön-ottoa ja hyödyntämistä. On hyvä huomata, että avoin lähdekoodi voi koskea muutakin kuin ”järjestelmiä”, esimerkkeinä tilastodatan ja julkaisujen tekoon soveltuvat R-ohjelmointikieli ja RStudio sekä paikkatietojärjestelmäohjelmisto QGIS.

Jo tietojärjestelmien hankintavaiheessa avoimen datan ja avointen rajapintojen huomioiminen on tärkeää ja siksi datan hyödyntämiseen liittyvää hankintaosaamista tulisi kehittää. Toimittajalukkoja muodostuu helposti, jos avoimen datan tietomalleja ei ole saatavilla, metatiedot puuttuvat, ohjelmointirajapintaan ei ole hyödyntämisesimerkkejä tai dokumentaatio puuttuu.

4.4.4 Osaamisen kehittäminen ja parhaiden käytäntöjen jakaminen

Julkisen avoimen datan osaaminen vaatii jatkuvaa kehittymistä ja kouluttautumista. Tietoyhteiskunnan kehitys muuttaa jatkuvasti tapoja hyödyntää avointa dataa. Datataito ei ole vain tekninen taito, eikä kosketa vain avointa dataa. Osaamisen tulee olla kokonaisvaltaista, ei vain ohjelmistorajapintoihin tai tiedonkäsittelyyn liittyvää. Julkishallinnon käynnistymässä oleva data- ja analytiikkaosaajien verkosto on hyvä esimerkki uudenlaisesta osaamisverkostosta. Ohjelmointi- ja datankäyttöaiheiset videoblogit ja screencast-videot (myös kaupalliset) ovat merkittävä ilmiö internetissä ja IT-maailmassa. Niillä voitaisiin kustannustehokkaasti ratkoa ongelmia ja tukea tehokasta ja oikeanlaista käyttöä.

Avoimen datan opas on julkaistu Avoindata.fi-palvelussa 7.4.2016 jatkona valtiovarainministeriön avoimen tiedon ohjelmalle. Avoimen datan opas ohjeistaa datan avaamisessa ja jakaa datan avaamisen hyviä käytäntöjä julkiselle hallinnolle ja muille kiinnostuneille tahoille. Opasta on ideoitu ja kehitetty käyttäjien palautteen perusteella yhdessä mm. julkishallinnon dataa avaavien virastojen, Suomen Kuntaliiton, Open Knowledge Finland ry:n ja Suomen

²¹ <http://www.jhs-suositukset.fi/suomi/jhs169>

²² <https://github.com/fmidev>

Esimerkki: Julkishallinnon datan avaajan mestarikurssi

Open Knowledge Finland ry on yhteistyössä Finnish Consulting Group Oy:n kanssa järjestänyt vuonna 2016 kaksi massiivista verkkokurssia (MOOC – Massive Open Online Course) nimeltään ”Datan avaajan mestarikurssi julkiselle hallinnolle”.

Kevään ja syksyn kursseilla oli yhteensä 194 kurssin suorittanutta henkilöä 35 organisaatiosta. Kursseilla virkamiehet perehtyivät avoimen datan peruskäsitteisiin ja datan avaamisen konkreettisiin toimiin. Osallistujat laativat omalle organisaatiolleen suunnitelman datan avaamisesta ja siitä tiedottamisesta. Monet kurssilaiset esittivät myös otteen avaamastaan datasta ja korjasivat sitä kurssin vetäjiltä ja muilta asiantuntijoilta saadun palautteen perusteella.

Kevään 2016 kurssin rahoitti opetus- ja kulttuuriministeriö ja syksyn 2016 kurssin Valtiokonttori. Kurssiaineiston kehittämisessä on tehty yhteistyötä valtiovarainministeriön, opetus- ja kulttuuriministeriön, liikenne- ja viestintäministeriön, Aalto-yliopiston, Maailmanpankkiryhmän sekä Helsinki Region Infosharen kanssa.

Kurssiaineistot ovat avoimesti saatavilla P2PU-palvelussa (Peer to Peer University).

kuuden suurimman kaupungin yhteisen 6Aika-hankkeen²³ kanssa. Avoimen datan opas oli haastateltaville pääosin tuttu ja osassa organisaatioissa käytetty ainakin jollain tasolla, vaikka siitä ei vahvaa mielipidettä ollutkaan. Liikennevirasto oli puolestaan luonut omat prosessinsa nimenomaan avoimen datan oppaassa ehdotetun prosessin mukaisesti.

Tyypillisesti haastatelluissa organisaatioissa oli ”avoimen datan osaajia”, eli henkilöitä, jotka ymmärtävät avoimen datan ”sillä tasolla, että voisivat pitää siitä esitelmän”, muutamista henkilöistä jopa kymmeneen. Osaamista tai sen puutetta ei kuitenkaan mainittu merkittäväksi ongelmaksi kovin monessa organisaatiossa. Työpajoissa nousi esille se, että verkkokoulutus ja videoluennot voivat olla tehokas tapa tavoittaa virkamiehet laajasti ja levittää parhaita käytäntöjä.

Haastatteluissa tuli esille myös mm. tilastojen ja datan lukutaito, jota tulisi opettaa niin oppilaitoksissa kuin työssäkin. Datan tuottajilla on jonkun verran huolia datan ”väärinkäytöstä” ja tulkinnasta. Usea haastateltavista toi esiin huolensa siitä, että heidän avoimena datana julkaisemaansa tietoa muunneltaisiin jollain tavalla väärin, eli osaamaton tai jopa manipulaatioon tähtäävä avoimen datan käsittely voisi heikentää avoimen datan tuottajan tai tietovarannon puolueettomuutta tai luotettavuutta. Useat haastateltavat kokivat, että he ovat jollain tasolla myös vastuussa siitä, että heidän avaamaansa dataa tulkitaan tai käytetään oikein.

4.4.5 Tehostuneet digitaaliset prosessit

Julkisen sektorin digitalisaatioon liittyviä ratkaisuja on kehitetty usein palvelu-, prosessi- ja virastokohtaisesti, eikä vanhoja rakenteita ole purettu. Uudet käyttöönotetut tietojärjestelmät eivät välttämättä pysty vaihtamaan tietoa (keskustelemaan) keskenään, eivätkä myöskään palvele riittävän hyvin kansalaisia, yrityksiä tai hallintoa itseään. Vertailtaessa esimerkiksi sähköisiä palveluita, tiedonhallintaa, tietoturva- tai arkistointia huomattiin, että myös kuntien väliset erot ovat suuria. Avoimen datan julkaisuun ja hyödyntämiseen vaikuttaa merkittävästi yleinen digitalisaatio- ja tietojärjestelmäkehitys. Avoin data ja keskitetyt tietovarannot voivat yksinkertaistaa julkisen sektorin palveluprosesseja ja avoin data mahdollistaa prosessien uudelleensuunnittelun asiakaslähtöisesti ja ketterästi. (Parviainen et al., 2017).

²³ <http://www.6aika.fi>

Erään haastateltavan mukaan viranomaisten käytössä ja sähköisesti saatavilla olevia tietoja kysytään asiakkailta edelleen turhaan paperilla. Syyksi tähän hän mainitsi asiakaslähtöisen toimintatavan puutteen, eli asiakas laitetaan keräämään paperit, koska on totuttu paperipohjaisiin prosesseihin. Muiksi syiksi hän arvioi, että osa julkisen sektorin toimijoista käyttää edelleen paperia, koska tietojärjestelmien ja ohjelmistorajapintojen hankintaan ei ole resursseja. Tiedon sähköinen käsittely on siis toteutunut vain osassa informaatioketjua ja tiedon elinkaarta. Toisinaan julkishallinto käyttää jalostettua avointa dataa myös kaupallisten palveluiden kautta. Haastateltava arvioi, että osa julkisen sektorin organisaatioista on valmis maksamaan tietojen helppokäyttöisyydestä yrityksille, vaikka samat tiedot olisivat viranomaisten käytössä ilmaiseksi.

Haastatteluissa nousi esiin se, että ministeriöt, virastot ja laitokset toimivat varsin erillään toisistaan, jolloin niiden välinen yhteistyö voi olla toisinaan työlästä ja hidasta tietojärjestelmien osalta. Tietojärjestelmien kirjava kokonaisuus on muodostunut jokaiseen virastoon ja ministeriöön vuosikymmenten aikana, jolloin yhteensopivuuden saavuttaminen on työlästä. Erään haastateltavan mukaan yhteistä kehittämistä riittävillä resursseilla pitäisi olla enemmän ja yksittäisten viranomaisten ”hajanainen puuhastelu” pitäisi lopettaa. Useat haastateltavat mainitsivat erilliset ohjelmointirajapinnat tai tietojärjestelmäintegraatiot, joita on rakennettu virastojen väliseen tiedonvaihtoon. Avoin data ja avoimet ohjelmointirajapinnat ovat edesauttaneet nimenomaan virastojen välistä yhteistyötä esim. Helsingin kaupungin toiminnassa.

Muita avoimuuteen liittyviä haastatteluissa mainittuja arvoja puolestaan olivat mm. oikeusvaltioperiaate, läpinäkyvyys, avoin lainvalmistelu ja hyvä hallintotapa.

4.5 Avoimen datan lainsäädäntö

Datan avaamiseen liittyy useita moniulotteisia lainsäädännöllisiä kysymyksiä. Tietovarantoihin ja perusrekistereihin liittyviä tai niitä sivuavia lakeja on vähintäänkin kymmeniä.

Tärkeimpiä yleisesti avoimeen dataan liittyviä lakeja ovat:

- Henkilötietolaki (523/1999 Henkilötietolaki)
- Julkisuuslaki (621/1999 Laki viranomaisten toiminnan julkisuudesta)
- Tietohallintolaki (634/2011 Laki julkisen hallinnon tietohallinnon ohjauksesta)
- Maksuperustelaki (150/1992 Valtion maksuperustelaki)
- KaPA-laki (571/2016 Laki hallinnon yhteisistä sähköisen asioinnin tukipalveluista).

Näiden yleislakien lisäksi virastoilla ja laitoksilla on myös niiden lakisääteisiä toimintoja ohjaavia lakeja, joilla ohjataan (perus)rekistereiden, tietopalveluiden tai tietovarantojen toimintaa²⁴. Yksistään henkilötiedon käyttöä tai tietosuojaa säädelään yli 600:ssa eri laissa ja asetuksessa (Nurmi, 2017).

Haastatteluissa ja työpajoissa nostettiin esille nykyiseen lainsäädäntöön, sääntelyn muuttuvaan kenttään (mm. EU:n uusi tietosuoja-asetus²⁵) sekä sääntelyn tulkintaan liittyvät haasteet, jotka hidastavat tietovarantojen avaamista ja hyödyntämistä. Lainsäädännölliset huomiot

²⁴ Esimerkiksi laki kiinteistöjärjestelmästä ja siitä tuotettavasti tietopalvelusta, kiinteistörekisterilaki, laki ja asetus väestötietojärjestelmästä, tilastolaki, geenivarantolaki, kalastuslaki ja metsälaki. Yleisiä eurooppalaisia säädöksiä ovat puolestaan mm. PSI-direktiivi, INSPIRE-direktiivi, Euroopan tilastoasetus sekä vuonna 2018 voimaan tuleva EU-tietosuoja-asetus.

²⁵ <http://eur-lex.europa.eu/eli/reg/2016/679/oj>

liittyvät mm. tiedon julkisuuteen, henkilötietoon ja tietosuojaan, tiedon maksullisuuteen ja avoimiin lisensseihin.

4.5.1 Avoimet lisenssit tiedon hyödyntämisen mahdollistajana

Avoimen datan lisenssin tarkoituksena on määritellä käyttöluva sille, miten dataa voi hyödyntää. Käyttöluvasta on annettu julkisen hallinnon suositus, eli JHS-suositus 189 – avoimen tietoaineiston käyttöluva, joka suosittelee uudelleenkäytön oikeudet julkishallinnon avoimiin tietoaineistoihin myönnettäväksi Creative Commons -lisenssiperheeseen kuuluvalla avoimella käyttöluvalla. Käyttöluvaksi suositellaan Creative Commons Nimeä 4.0 (CC BY 4.0), jolloin aineistoa saa vapaasti käyttää, kunhan nimeää sen tuottajan. Jos datan yksilöiminen tai tuottajan tietojen ilmoittaminen ei ole tarpeellista, voidaan käyttää CC0-käyttöilupaa, mikä tarkoittaa, että julkaisija luopuu kaikista yksinoikeuksistaan, joista hän lain mukaan voi luopua. Esimerkiksi tietoaineiston metatiedot tulisi julkaista aina CC0-käyttöluvalla.

Tutkimuksessa tuli esille, että lähes kaikkien organisaatioiden avoin data on julkaistu CC 4.0 BY -lisenssin mukaisesti. Toisinaan kuitenkin avoimen datan hyödyntäjät unohtavat mainita avoimen datan lähteen, mikä tietyllä tavalla haittaa avoimen datan vaikuttavuuden selvittämistä. Lähde tulee todennäköisemmin mainituksi, jos hyödyntäjälle annetaan tarkka ohje, kuinka lähde tulisi mainita, kuten Liikennevirasto tekee²⁶.

4.5.2 Haasteena vaihtelevat lain tulkinnat

Haastatteluiden perusteella erityisesti henkilötietolain, julkisuuslain ja maksuperustelain kohdalla oli haasteita liittyen mm. lakien tulkintaan ja soveltamisohjeisiin. Tulkintakysymykset vaativat harkintaa ja saattavat kuormittaa perusrekistereiden ylläpitäjiä toisinaan paljonkin.

Julkisuuslaki (621/1999) on yleissäännös, joka käsittelee viranomaisten hallussa olevien asiakirjojen julkisuutta, salaamista ja julkistamista. Perussäännön mukaan asiakirja on julkinen, ellei sen salaamiseen ole laissa erikseen mainittua syytä. Samaan aikaan Henkilötietolaki suojaa yksityisyyttä ja rajoittaa henkilötiedon käsittelyä ja julkaisua. Julkisuuslakia tulkitaan monin eri tavoin ja sen soveltaminen ehkä tarvitsisi ohjeistusta sekä lain soveltamiseen ja kouluttamiseen keskittyvän tahon. Yhdistyneen kuningaskunnan *Information Commission*²⁷ on kansainvälinen esimerkki viranomaisesta, joka vastaa sekä tiedon avoimuudesta että tietosuojasta.

Maksuperustelain mukaan monissa tapauksissa julkishallinnon tuottaman datan käytöstä voidaan peritä korvaus (Poikola et al., 2010). Maksuperustelaki mahdollistaa myös datan julkaisemisen tai luovuttamisen veloitusetta. Haastattelujen perusteella lain tulkinnat ja soveltamiskäytännöt vaihtelevat: osa haastatelluista näki laissa joustoa, jolloin sen tulkinta ei estä avoimen datan julkaisua. Osa taas tulkitsi lakia siten, että maksuperustelain seurauksena jotain dataa ei voida avata saataville maksutta. Yhtenäiset käytännöt vaikuttivat puuttuvan.

Tehokkaan datan hyödyntämisen kannalta pitäisi vähentää säädösten ja niiden tulkintojen sekä sovellusohjeiden määrää ja koota yhteen parhaita käytäntöjä, joita levitettäisiin koordinoitusti ministeriöihin, virastoihin, laitoksiin ja kuntiin. Osana hallituksen digitalisaatio-kärkihanketta²⁸ onkin valmisteilla uusi tiedonhallintalaki, jossa myös tietovarantojen avaamiseen

²⁶ http://www.liikennevirasto.fi/avoindata/kayttoehdot#.WK80MG_c7IU

²⁷ <https://ico.org.uk/>

²⁸ <http://valtioneuvosto.fi/hallitusohjelman-toteutus/digitalisaatio/karkihanke1>

liittyviä seikkoja tultaneen tarkentamaan. Liikennekaari-hankkeessa puolestaan hajanaista liikenteen rekisterilainsäädäntöä kootaan yhteen lakiin. Uudistuksen tavoite on nykyistä huomattavasti tehokkaampi tiedon hyödyntäminen, rekisteröidyn oikeuksien toteuttaminen, omadatan edistäminen sekä yhteentoimivuuden parantaminen.

4.5.3 Henkilötiedon ja avoimen datan välinen suhde kaipaa selvennyksiä

Henkilötietolaki on henkilötietojen käsittelyn peruslaki, joka on säädetty toteuttamaan yksityiselämän suojaa sekä muita yksityisyyden suojaa turvaavia perusoikeuksia henkilötietojen käsittelyssä sekä edistämään hyvää tietojenkäsittelytapaa. Henkilötietolain noudattamista valvoo ja ohjeistaa tietosuojavaltuutettu.

Henkilötietolain 3 §:n mukaan henkilötiedolla tarkoitetaan kaikenlaisia luonnollista henkilöä taikka hänen ominaisuuksiaan tai elinolosuhteitaan kuvaavia merkintöjä, jotka voidaan tunnistaa häntä tai hänen perhettään tai hänen kanssaan yhteisessä taloudessa eläviä koskeviksi. Vaikka jokin henkilötieto olisikin julkisuuslain nojalla julkinen tieto, ei se tarkoita sitä, että tieto voitaisiin saattaa vapaasti uudelleen käytettäväksi.

Haastatteluissa ja työpajoissa mainittuja henkilötiedon ja avoimen datan hyödyntämiseen liittyviä epäselviä tapauksia olivat esimerkiksi verotustiedot, ostolaskut sekä yritystiedot. Moni haastateltava nosti esiin henkilötietolain ja julkisuuslain välisen paikoitellen vaikean tulkinnan. Virastoilla voi olla erilaisia ohjeita ja käytäntöjä sen suhteen, miten avoimen datan ja henkilötiedon kanssa toimitaan, ja siten samankaltaisten ongelmien kanssa painitaan eri paikoissa. Useimmilla haastatelluilla organisaatioilla oli jo alkanut tai alkamassa projekti EU:n tietosuoja-asetukseen valmistautumisesta ja siten tietosuoja-asetuksen uudet velvoitteet olivat pääosin tuttuja. Tuleva tietosuoja-asetus nähtiin hyvänä asiana siitä näkökulmasta, että lait saadaan paremmin vastaamaan nykyisiä tarpeita.

Verotuksen toimittamiseen käytettävät tiedot ovat lähtökohtaisesti aina salassa pidettäviä. Tätä perustellaan sillä, että Verohallinnon on saatava tietoja verotuksen kohteena olevan omalla ilmoituksella, jolloin ilmoitushalukkuus voisi kärsiä, mikäli tiedot olisivat täysin avoimia. Verotuksen lopputulos on julkinen ja selailtavissa Verohallinnon tiloissa, joskaan yksityishenkilöiden tietoja ei julkaista avoimena datana. Yhteisöjen verotus sen sijaan löytyy avoimena aineistona.

Lähtökohtaisesti yrityksiä koskevan tiedon ja henkilötiedon rajan pitäisi olla selkeä. Yritystietojen tulisi olla julkisia ja avoimia esimerkiksi kuntien ja valtion ostolaskudatassa. Esimerkiksi toiminimien kohdalla tilanne on kuitenkin tulkinnanvarainen, sillä yrityksen nimi tai osoite voi sisältää henkilötietoa. Tämä ristiriita on hidastanut esimerkiksi kuntien ostolaskujen avaamista ja kunnissa on toisistaan poikkeavia käytäntöjä sen suhteen, sisällytetäänkö toiminimet avattavaan ostolaskudataan vai ei.

Yrityksistä sekä niiden taloustiedoista ja omistajista on Suomessa melko vähän tietoa saatavilla. Kansainvälisessä Global Open Data Index -vertailussa Suomen kokonaissijoitus laskee, koska yritystietoja ei saa ladattua kokonaisuudessaan PRH:n sivuilta. Esimerkiksi Yhdistyneellä kuningaskunnalla on yritysten omistajia luetteloiva People with Significant Control Register²⁹ ja kansainvälinen Open Ownership³⁰ -hanke puolestaan on saanut taakseen toistakymmentä maata, tavoitteena kerätä kattavaa tietoa tosiasiallisista edunsaajista. Sekä

²⁹ <https://www.gov.uk/government/publications/guidance-to-the-people-with-significant-control-requirements-for-companies-and-limited-liability-partnerships>

³⁰ <http://openownership.org/>

haastatteluissa että työpajoissa tuotiin esiin yritysten tiedon avaamisen hyödyntämismahdollisuudet (mm. verojen läpinäkyvyys, harmaan talouden torjunta, kilpailija-analyysi ja lisääntynyt kilpailu).

EU:n yleinen tietosuojasetus (General Data Protection Regulation, GDPR) on uusi kaikkia EU-maita koskeva tietosuojasetus, joka astuu voimaan 25.5.2018. Tietosuojasetuksessa on laaja kansallinen liikkumavara ja sen vaikutus Suomen lainsäädäntöön tarkentuu. EU:n uuden tietosuojasetuksen takia oikeusministeriö suosittelee omassa ohjeessaan tietolinjapäättökseen tekemistä osana asetukseen liittyvää valmistautumisprosessia (Talus & Autio 2017).

4.6 Tietoinfrastruktuuri ja teknologia

Tietotyö perustuu organisaatioiden tehokkaaseen tiedonhallintaan. Tiedonhallinta voidaan jäsentää kahteen osa-alueeseen, hiljaisen ja eksplisiittisen tiedon hallinnaksi. Kustannustehokkuutta saavutetaan, jos molemmat tiedon osa-alueet huomioidaan.

Tietotekniikan avulla käsitellään ainoastaan eksplisiittistä tietoa, eli kirjallisessa muodossa olevaa tietoa, jota voidaan tallentaa ja siirrellä helposti. Luonnolliset kielet, tietomallit, tietokannat, ohjelmointikielet, tietorakenteet ja matemaattiset ilmaisut ovat esimerkkejä eksplisiittisen tiedon ilmaisu- ja siirtövälineistä. Eksplisiittisen tiedon kuvaillaan usein olevan tieteellistä, dokumentoitua ja tallennettua ja se ilmaistaan sanoina ja numeroina esimerkiksi tietovarantojen, dokumenttien, raporttien ja muistiinpanojen avulla.

Hiljainen tieto on kokemuksen kautta kertynyttä tietämystä, osin tiedostettua ja tiedostamattomaa. Hiljaista tietoa voidaan kuvata esimerkiksi intuition ja osaamisena. Hiljaisen tiedon kuvaaminen sanallisessa muodossa on vaikeaa ja sen siirtäminen henkilöltä toiselle voi olla haasteellista. Organisaation tiedosta suurin osa on hiljaista tietoa (muun muassa työkokemus suurelta osin).

Tietoinfrastruktuuri koostuu muun muassa tietovarannoista, niitä hallinnoivista ja ylläpitävistä organisaatioista sekä ohjeistuksista ja käytännöistä, joiden avulla tietoa käytetään ja hallinnoidaan. Tietoinfrastruktuuri tulisi nähdä vähintään yhtä tärkeänä kuin fyysinen infrastruktuuri (esimerkiksi tiet, rakennukset, tehtaot, sähköverkko ja tietoliikenneverkko). Tietojohtamista tulisi soveltaa kokonaisvaltaisesti. Esimerkiksi Ruotsilla on ollut digitalisaatiosta vastaava ministeri (ruotsiksi: IT-minister) vuodesta 2016 lähtien.

Tiedon merkitys osana toimintaa on ymmärretty hyvin Trafissa. Trafi määrittelee tiedon olevan viides liikennemuoto tieliikenteen, ilmailun, merenkulun ja raideliikenteen ohella. Tiedon nähdään olevan muita liikennemuotoja läpileikkaava teema ja tulevaisuudessa lähes kaikkien liikkumiseen liittyvien toimintojen polttoaine ja ohjaaja (Trafi, 2016).

Kaikki digitaalisessa muodossa oleva organisaation tieto on eksplisiittistä, mutta sen hyödyntämiseen tarvitaan usein hiljaista tietoa. Avoimen tiedon tulkitseminen ja hyödyntäminen hyödyntävät eksplisiittisen tiedon lisäksi aina hiljaista tietoa. Avoimen datan kustannustehokkaan hyödyntämisen kannalta on tärkeää tiedostaa näiden molempien olemassaolo ja toisiaan täydentävä yhteisvaikutus.

Taulukko 4.4 Selvityksen organisaatioiden avoimen datan tarjonta

Organisaatio	Avoimen datan sivu(t)	Avoimen datan julkaisu-tapa (rajapinta, tiedostot, katselu- ja latauspalvelu)	Avointen aineistojen lisenssi	Rekisteröityminen	Tietoaineistoja Avoindata.fi:ssä
Ilmatieteen laitos	https://ilmatieteenlaitos.fi/avoin-data	Rajapinta, tiedostot, katselu- ja latauspalvelu	JHS 189 / CC BY 4.0	Taustatiedot käyttäjistä	29
Kela	http://www.kela.fi/kelasto	Tiedostot, katselu- ja latauspalvelu	JHS 189 / CC BY 4.0	Ei	0
Liikennevirasto	http://www.liikennevirasto.fi/avoindata	Rajapinta, tiedostot, katselu- ja latauspalvelu	JHS 189 / CC BY 4.0; joitain rajoituksia esim. ei navigointiin	Sähköposti tiedosto-lataukseen	141
Liikenteen turvallisuusvirasto (Trafi)	http://www.trafi.fi/tietopalvelut/avoin_data	Tiedostot	JHS 189 / CC BY 4.0	Ei	3
Luonnonvarakeskus	http://stat.luke.fi/	Rajapinta, katselu- ja latauspalvelu	JHS 189 / CC BY 4.0	Ei	2
Maanmittauslaitos (MML)	http://www.maanmittauslaitos.fi/aineistot-palvelut/maastotiedot	Rajapinta, tiedostot, katselu- ja latauspalvelu	JHS 189 / CC BY 4.0	Osaan palveluista yhteystiedot	62
Patentti- ja rekisterihallitus (PRH)	http://avoindata.prh.fi/	Rajapinta, katselupalvelu	JHS 189 / CC BY 4.0	Ei	1
Suomen ympäristökeskus (SYKE)	http://www.syke.fi/avoindata	Rajapinta, tiedostot, katselu- ja latauspalvelu	JHS 189 / CC BY 4.0	Ympäristöpalveluihin sähköposti ja taustatiedot	91
Terveystieteiden tutkimuskeskus (THL)	https://www.thl.fi/avoindata	Rajapinta, tiedostot, katselu- ja latauspalvelu	JHS 189 / CC BY 4.0	Kirjautumalla laajempi toiminnallisuus joissain palveluissa	4
Tilastokeskus	https://www.stat.fi/org/avoindata/index.html	Rajapinta, tiedostot, katselu- ja latauspalvelu	Oma (JHS 189 / CC BY 4.0 tulossa)	Osaan palveluista sähköposti ja taustatiedot	21
Väestörekisterikeskus (VRK)	(avoindata.fi:ssä)	Tiedostot	JHS 189 / CC BY 4.0	Ei	4
Verohallinto	https://www.vero.fi/fi-FI/Avoin_data(29357)	Tiedostot	JHS 189 / CC BY 4.0	Ei	2

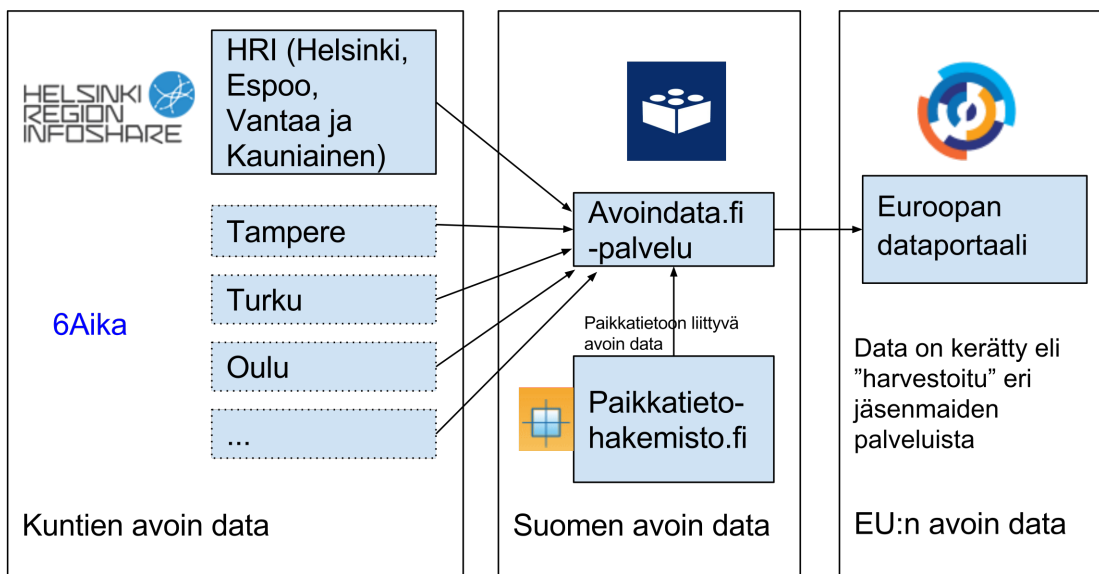
Taulukossa 4.4 on kuvattu organisaatioiden avoimen datan tarjontaa. Lähes kaikilla organisaatioilla on oma avoimen datan jakelukanava (tai kanavia). Sen lisäksi tietoa jaetaan Avoindata.fi -portaalissa. Yksi organisaatioista jakaa avointa dataa vain omien sivujensa kautta ja yksi vain Avoindata.fi -palvelun kautta. Noin puolet organisaatioista kerää käyttäjäkohtaista tietoa ja kaikki paitsi yksi käyttää JHS 189 -suosituksen mukaista lisenssiä.

4.6.1 Avoindata.fi ja dataportaalit

Avointa tietoa jaetaan pääsääntöisesti kolmella tavalla: tiedostoina, katselu- ja latauspalvelun kautta ja ohjelmointirajapinnan välityksellä. Avoimen datan portaalit (tai datakatalogit) ovat keskeinen osa tietoinfrastruktuuria, jonka avulla datan hyödyntäjät löytävät julkishallinnon eri organisaatioiden avointa dataa yhdestä paikasta. Maailmanlaajuisesti on olemassa yli 2 600 avoimen datan portaalit (OpenDataSoft, 2017). Usein kaupungeilla ja mailla on oma dataportaalinsa.

Avoindata.fi-palvelu on Suomen julkishallinnon avoimen datan ja tiedon yhteensopivuuden portaalit, joka julkaistiin osana valtiovarainministeriön avoimen tiedon ohjelmaa (2013–2015) syksyllä 2014. Palvelussa voi etsiä muiden julkaisemia tietoaineistoja sekä julkaista ja hallinnoida avoimia tietoaineistoja. Palveluun siirretään (harvestoidaan) automaattisesti metatiedot myös pääkaupunkiseudun yhteisestä avoimen datan portaalista (Helsinki Region Infoshare,

Kuvio 4.1 Avoimen datan kerääminen ja siirtäminen kuntatasolta EU-tasolle



Havainnollistava kuvio.

HRI³¹) ja Maanmittauslaitoksen ylläpitämästä paikkatietohakemistosta³² (paikkatietoihin liittyvä avoin data).

Avoindata.fi, Euroopan avoimen datan portaali³³ ja HRI-palvelu toimivat avoimen lähdekoodin CKAN³⁴-ohjelmiston päällä. CKAN-alustalle tallennettua dataa voi hyödyntää koneluettavan ohjelmointirajapinnan tai verkkosivuston kautta.

Avoindata.fi-palvelussa on yli 1 600 tietoaainetta (22.2.2017). Palveluun on yhdistetty erillinen yhteentoimivuus.fi-palvelu, joka sisältää 321 tietoaainetta yhteentoimivuuden ja kokonaisarkkitehtuurin kuvauksia ja ohjeita sekä välittää ajankohtaista tietoa kehittämistyöstä ja koulutuksesta. Suunnittelutiedon jakamisella pyritään uudelleenkäyttöön ja päällekkäisen työn vähentämiseen.

Paikkatietohakemisto on Maanmittauslaitoksen ylläpitämä julkinen ja avoin verkkosivusto, jonne on koottu paikkatietoja koskevaa metadattaa, esimerkiksi INSPIRE-direktiivin määrittelemät aineistot luokiteltuna aihealueittain. Esimerkiksi Luonnonvarakeskuksen ja Suomen ympäristökeskuksen avoin data on suurelta osin paikkatietosidonnaista. Paikkatietoihin sidottua avointa tietoa on lisäksi Paikkatietoikkuna.fi-palvelussa.

Euroopanlaajuinen avoimen datan portaali kokoaa jäsenmaista avointa dataa keräämällä (harvestoimalla) tietoa jäsenmaiden avoimen datan portaaleista. Portaalissa on yli 600 000 aineistoa 34:stä maasta 18:lla kielellä (Sasse et al., 2017, s. 7). Euroopan komissio on määrittellyt portaalin tehtävän seuraavasti: ”EU:n avoimen datan portaali tarjoaa keskitetyn pääsyn EU:n toimielinten ja laitosten tuottamaan jatkuvasti kasvavaan datavarantoon” (Euroopan komissio 2017). Suomesta kerätään automaattisesti tietoaainetta Euroopan avoimen datan

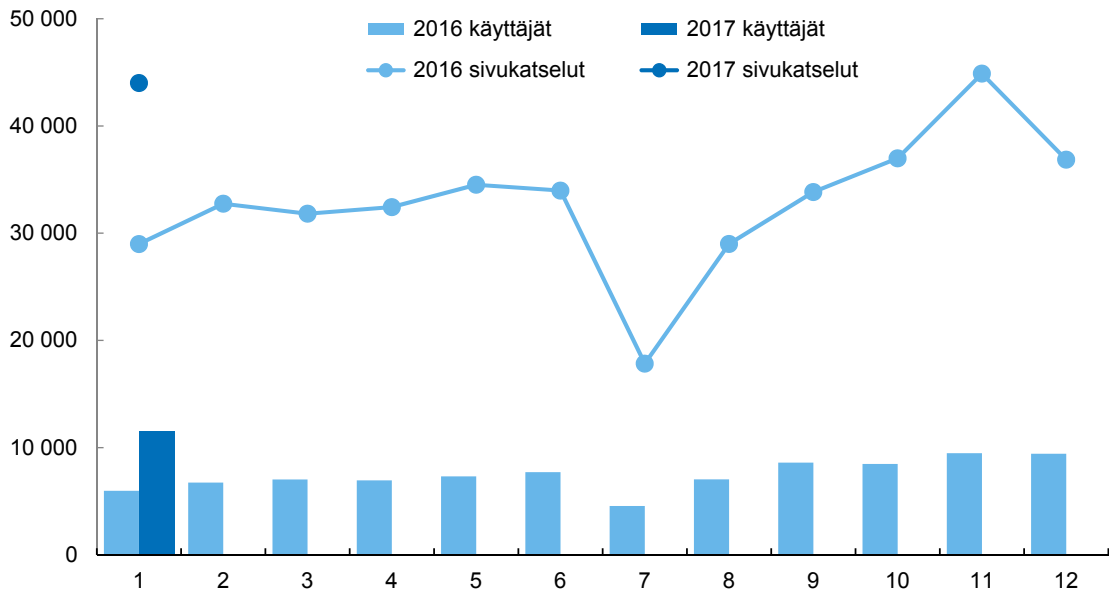
³¹ <http://www.hri.fi/fi/>

³² www.paikkatietohakemisto.fi

³³ <https://www.europeandataportal.eu>

³⁴ <http://www.ckan.org>

Kuvio 4.2 Avoindata.fi-palvelun käyttötilastot



Lähde: www.avoindata.fi. Kuviossa pystyakselilla ovat kävijämäärät ja vaakakselilla kuukaudet.

portaaliin Avoindata.fi-palvelun kautta. Suomesta siirrettävät tietoaineistot ja niiden metadata käännetään koneellisesti englanninkielisiksi, jonka seurauksena käännösten laatu ja ymmärrettävyys ovat heikkoja. Osa materiaalista jää suomenkieliseksi.

Avoindata.fi-palvelun käyttöä (kuvio 4.2) seurataan tilastoimalla istuntoja, käyttäjiä ja sivukatseluita. Tilastot kuitenkin ovat vain suuntaa antava mittari datan käytölle ja heikko indikaattori vaikuttavuudelle. Helsingin kaupungin julkaisemaa avointa dataa esimerkiksi jaetaan Hri.fi:ssä, Avoindata.fi:ssä ja Euroopan dataportaalissa, joten yhden dataportaalin käyttäjämäärät eivät ole luotettava datan käyttöä kuvaavia mittareita.

Avoindata.fi-palvelu oli tuttu yleisellä tasolla kaikille haastatelluille, mutta sen tarkempi tunnettuus ja käyttö vaihteli. Haastateltavista vain yhdellä organisaatiolla ei ollut yhtään tietoaineistoa tai aineistokuvausta Avoindata.fi-palvelussa. Tiedon löytämisen haasteet nousivat esiin myös työpajoissa. Voidaan sanoa, että tiedon löytäminen ei ole selkeä kokonaisuus, eli esimerkiksi virastojen ja laitosten omista dataportaleista ja verkkosivuilta puuttuu usein linkityksiä Avoindata.fi-palveluun. Vastaavasti Avoindata.fi:stä löytyy vain osa datasta ja esimerkiksi toimimattomia linkkejä tietoaineistoihin. Riippuen käyttötapauksesta ja hyödyntäjän kokemuksesta, luonteva paikka etsiä tarvittavaa dataa voi olla joko viraston tai laitoksen oma palvelu tai yleisempi datakatalogi, kuten Avoindata.fi-palvelu. Koska datan avaajilla on vaihtelevia tapoja ja kanavia julkaista avointa dataa eri kanavissa, datan hyödyntäjän pitää usein osata etsiä oikeasta paikasta löytääkseen haluamaansa dataa.

4.6.2 Katselu- ja latauspalvelut ja ohjelmointirajapinnat

Katselu- ja latauspalvelu on verkkosivu tai -palvelu, jonka kautta voi ladata ("kopioida") tietyt avoimet tiedostot paikallisesti toiselle koneelle tai palvelimelle (ks. myös rajapintapalvelu), kunhan on valittu kriteerit, esimerkiksi tietokantataulukot tai maantieteellinen alue. Esimerkiksi

Tilastokeskuksen PX-Web -latauspalvelussa³⁵ voi rajata ja ladata useita erilaisia tilastoaineistoja. Datat saa ladattua useassa eri tiedostoformaatissa (myös rakenteellisessa muodossa). Aineistot on luokiteltu ja järjestetty kansiomaiseen tietorakenteeseen. Ensin valitaan selaimella tai etsitään hakusanalla haluttu aineisto ja valitaan siitä ennalta määritetyillä ehdoilla haluttu osa. Lopuksi valitaan haluttu tiedostomuoto, jossa data halutaan ladata. Datassa käytettävä lisenssi on Tilastokeskuksen oma, joka poikkeaa hieman JHS 189 suosituksesta. Katselu ja latauspalvelut ovat hyödyllisiä, kun valittavaa aineistoa on potentiaalisesti paljon kuten tilastoja tai paikkatietoaineistoja. Katselu ja latauspalveluita voi tyypillisesti käyttää sekä käyttöliittymän että rajapinnan kautta.

Ohjelmointirajapinta on ohjelmiston osa, jonka avulla optimoidaan tietoturvallista tiedonsiirtoa eri ohjelmistojen välillä (Stowe, 2015, s. 1–2). Ohjelmointirajapinnan tai latauspalvelun avulla avoimena datana julkaistuista kokonaistiedoista voidaan ladata vain tarvittu osakokonaisuus määrittelemällä tietojoukkoa rajaavia hakuehtoja. Ohjelmointirajapinta on tehokas tapa jakaa tietoa erityisesti silloin, kun tietomäärä on suuri, data on reaaliaikaista tai muuttuu usein.

Avoimen rajapinnan määritelmän³⁶ mukaan ohjelmointirajapintojen tulee olla käytettävissä ilman tunnistautumista tai vaihtoehtoisesti automaattisen rekisteröinnin avulla. Osa tietoväestön tarjoajista edellyttää rekisteröintiä ja kerää taustatietoja, esimerkiksi käyttäjän sähköpostiosoitteen. Tällöin ohjelmointirajapinnan hyödyntämiseen tarvittava avainkoodi lähetetään käyttäjän ilmoittamaan sähköpostiosoitteeseen ja jokainen rajapintaan tehtävä kysely toimii vain, jos se sisältää saadun yksilöllisen avainkoodin. Ymmärrys datan hyödyntäjien tarpeista voi lisääntyä, jos avoimen datan käyttäjiä voidaan tunnistaa jollain tasolla. Tällöin datan tuottajan on helpompaa esimerkiksi ilmoittaa muuttuneista ominaisuuksista ja käyttöehdoista tai muuten kerätä palautetta datan hyödyntäjiltä.

Tiedon jakelu ohjelmointirajapintojen kautta on tehokasta, mutta voi aiheuttaa kasvavia palvelin- ja tietoliikennekustannuksia kasvavan käytön myötä. Yksi tapa datan julkaisijalle hallita ja ennakoita kustannuksia on ohjelmointirajapinnan käytön rajoittaminen. Osa tietoa-aineistojen tarjoajista rajoittaa rajapintojen käyttöä esimerkiksi kuorman ja siten kustannuksien hallitsemiseksi tai lisätulojen hankkimiseksi. Esimerkiksi Ilmatieteen avoimen datan palvelun käyttörajoitukset ovat 1) latauspalveluun enintään 20 000 kyselyä vuorokaudessa, 2) katselupalveluun 10 000 kyselyä vuorokaudessa, 3) molempiin palveluihin 600 kyselyä viidessä minuutissa (Ilmatieteen laitos, 2017). Mikäli käyttö ylittää nämä määrät, käytöstä joutuu maksamaan omakustannushinnan. Ohjelmointirajapintaa pääsee käyttämään ainoastaan ennakkorekisteröitymällä sen käyttäjäksi.

Patentti- ja rekisterihallituksen lähestymistapa avoimen datan rajapintoihin on toisenlainen. Ohjelmointirajapinnan käyttö ei vaadi rekisteröitymistä, mutta palvelun käyttö on kokonaiskapasiteetiltan rajoitettua. Ohjelmointirajapinnan maksimikapasiteetti, joka jakautuu käyttäjien kesken, on 300 hakua minuutissa (Patentti- ja rekisterihallitus, 2017). On sattumanvaraista, missä vaiheessa tuo 300 kyselyä tulee täyteen. Nykyinen PRH:n avoimen datan ohjelmointirajapinta rajoittaa merkittävästi saatavuutta, sillä on mahdollista, että esimerkiksi yksi käyttäjä käyttää rajapinnan kapasiteetin kokonaan. PRH:n avoimen datan ohjelmointirajapinnan käyttö on suurta. Vuonna 2015 ohjelmointirajapintaan tehtiin 53 miljoonaa hakua. Esimerkiksi YTJ-järjestelmään tehtiin samaan aikaan 20,9 miljoonaa hakua. (PRH, 2016, s. 12).

³⁵ <http://pxnet2.stat.fi/PXWeb/pxweb/fi/StatFin/>
http://www.liikennevirasto.fi/avoindata/katselu-lataus#.WLK_KW_c7IU
<http://www.maanmittauslaitos.fi/aineistot-palvelut/latauspalvelut/avoimien-aineistojen-tiedostopalvelu>
<https://www.Avoindata.fi/data/fi/dataset/syken-inspire-latauspalvelu-bet>

³⁶ <http://avoimirajapinta.fi/>

Suuri ohjelmointirajapinnan käyttömäärä ei välttämättä kerro datan laajasta käytöstä, vaan suuri käyttömäärä voi johtua myös ohjelmointirajapinnan epäluotettavuudesta. Silloin varmuuden vuoksi tiedon hyödyntäjät lataavat ja tallentavat tietoja toistuvasti varmuuden vuoksi omaan tietovarantonsa. Huonosti toteutettu rajapinta ei palvele ketään. Tiedon hyödyntäjältä se kuluttaa aikaa, vaatii ylimääräistä työtä ja tiedon tallennustilaa, ja tiedon avaajalle se voi tuottaa esimerkiksi imago-tappioita ja ohjelmointirajapinnan ylimääräistä kuormitusta. Lisäksi ohjelmointirajapintojen julkaiseminen ei vielä riitä siihen, että niitä käytettäisiin. Esimerkiksi ohjelmointirajapinta, jolla ei ole palvelulupausta, aktiivista tukea tai käytännön esimerkkejä voi jäädä hyödyntämättä. Ohjelmointirajapinta on tiedostojakelua vaativampi jakelumuoto siinä mielessä, että se voi vaatia enemmän tukea datan hyödyntäjille.

Avoimen datan jakeluun käytettäviä ohjelmointirajapintoja on haastatteluissa organisaatioissa käytössä monenlaisia ja rajapintakehitykseen panostetaan organisaatioissa. Tarkasti avoimen ohjelmointirajapinnan määritelmään verrattuna mikään avoimen datan jakeluun käytettävä ohjelmointirajapinta ei kuitenkaan ollut täysin avoin, sillä puutteita oli mm. palvelutasosopimuksissa, dokumentaatioissa, käyttöesimerkeissä ja -rajoituksissa.

Ensimmäisessä työpajassa tuli esille, että usea datan hyödyntäjä oli kokenut Ilmatieteen laitoksen ohjelmointirajapinnan hankalaksi käyttää. Toisessa työpajassa tuli esille ajatus siitä, että avoimen datan hyödyntäjille tarjottaisiin vapaaehtoinen mahdollisuus rekisteröityä aineiston hyödyntäjiksi. Hyödyntäjän on usein tarpeellista tietää ennakkolta avoimen dataan tai ohjelmointirajapintoihin tehtävistä muutoksista, jotta dataa hyödyntävän sovelluksen tai palvelun toiminta ei keskeydy tai häiriinny.

Haastattelujen perusteella nykyisiin avoimen datan ohjelmointirajapintoihin liittyi monenlaisia haasteita. Esimerkiksi Tilastokeskuksen ohjelmointirajapinnat ovat toisaalta EU-standardeihin pohjautuvia, mutta samalla osa käyttäjistä piti niiden hyödyntämistä monimutkaisena toteutuksen takia vaikeana. Lähtökohtaisesti EU-tasolla määritellyt ohjelmointirajapinnat ovat hyvä ajatus, koska ne parantavat skaalautuvuutta. Esimerkiksi avointa dataa hyödyntävän älypuhelinsovelluksen potentiaalisina markkinoina voivat olla Suomen lisäksi kaikki muutkin EU-maat.

4.6.3 Avoimen datan laadun parantaminen

Avoimessa datassa on paljon erilaisia datan hyödynnettävyyteen vaikuttavia piirteitä kuten datan ajantasaisuus ja yhteensopivuus tiedostomuotojen, tietomallien ja metadatan kautta.

On olemassa erilaisia tapoja ohjata tietovarantojen tietojen päivittämistä. Väestötietojärjestelmä (VTJ) rakentuu ilmoitusvelvollisuuden varaan ja tietojen ilmoittaminen rekisteriin on ilmaista. VTJ:n ajantasaisuus onkin erinomainen ja tarkasti säädelty. Tietojen päivittäminen Patentti- ja rekisterihallituksen yhdistysrekisteriin perustuu puolestaan vapaaehtoiseen ja pääosin maksulliseen ilmoittamiseen, joka voi vaikuttaa tietojen päivittämiseen ajantasaisiksi. Yhdistysrekisterissä on 40 000 yhdistystä, joiden tietoja ei ole päivitetty vuoden 1995 jälkeen³⁷. Tämä on noin 30 % kaikista yhdistyksistä, kun Suomessa on yhteensä noin 135 000 yhdistystä³⁸. Näin ollen tietojen oikeellisuuden ja julkisen avoimen datan oikeellisuuden ja laadukkuuden kannalta on parempi, jos tiedot kerätään automaattisesti, niiden ilmoittaminen on ilmaista ja ajantasaisia tietoja veloitetaan lainsäädännössä. Vanhentunut tai muuten heikkolaatuinen raakadata ei muutu hyväksi osana datan avaamisen prosessia.

³⁷ https://www.prh.fi/fi/uutislistaus/2016/P_8973.html

³⁸ <http://www.hs.fi/kotimaa/art-2000002915151.html>

Taulukko 4.5 Kahden perusrekisterin erilainen lähestymistapa ylläpitää tietoja (tiedonhallintaprosessi)

	Ilmoittaminen	Hakeminen	Ajantasaisuus	Ilmoitusvelvollisuus
Väestötietojärjestelmä	Ilmaista	Maksaa	Hyvä	Ilmoitettava
Yhdistysrekisteri	Pääosin maksullista	Ilmaista	> 35 % vanhoja (ei päivitetty vuoden 1995 jälkeen)	Tulkinnanvaraista

Avoimen datan laadukkuuteen kuuluu julkaiseminen avoimessa ja yleisesti tunnetussa tiedostoformaateissa sekä kattava metadata, jotka edistävät datan jatkokäyttöä ja automaattista hyödyntämistä. Datan laadukkuutta voidaan arvioida eri mittaristojen avulla. Yksi mittaristo on World Wide Webin kehittäjän Tim Berners-Leen vuonna 2009 esittelemä avoimen datan viisiasteinen tähtiluokitus (Berners-Lee, 2009). Avoimeen dataan vaaditaan vähintään kolme tähteä. Tasoilla neljä ja viisi painotetaan linkitetyn datan ja semanttisen webin piirteitä. Vain neljä tietoaineistoa saa täydet viisi tähteä tällä mittaristolla Avoindata.fi-palvelussa (20.2.2017).

Taulukko 4.6 Viiden tähden arviointimalli avoimen datan laadulle

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

Vaihtoehtoinen ja uudempi menetelmä datan uudelleenkäytettävyyden parantamiseksi on kitkaton datapaketti (frictionless data³⁹). Siinä data on nimensä mukaisesti paketissa, jonka sisällä on eroteltu data(tiedosto) ja metadata(tiedosto) erikseen. Datatiedosto voi olla esimerkiksi CSV-formaatissa ja metadata tiedosto JSON-formaatissa.

Avoimen datan puutteelliset tai monitulkinnalliset metatiedot ovat yksi suurimpia ongelmia, mitä avoimeen dataan liittyy (Economist, 2015). Tärkeää olisi tunnisteiden, datan ja metadatan hallinta. Osa metatietoja on datan tietomallin julkaiseminen. Tutkimusdatan laadukkaalle julkaisulle on puolestaan FAIR-periaatteet⁴⁰. FAIR on lyhenne sanoista Findable, Accessible, Interoperable ja Re-usable ja suomeksi voitaisiin puhua tutkimusaineistojen löydettävyydestä, saavutettavuudesta, yhteentoimivuudesta ja uudelleenkäytettävyydestä.

Avoimen datan hyödyntämistä helpottaa, jos datan mukana julkaistaan myös sen tietomalli. Tietomalli on abstrakti malli datan loogisesta rakenteesta ja esimerkiksi sen sisältämästä metadatasta. Ilman kunnollista tietomallia avoin data voi olla monitulkintaista ja vaikeasti hyödynnettävissä. Tietomallit puuttuvat usein avoimesta datasta. Hyvä käytännön esimerkki tästä on Helsingin, Espoon ja Vantaan julkaisema kaupunkien ostolaskudata. Datassa esimerkiksi tilikartat, laskutus sopimukset, kustannuspaikkajako ovat erilaisia kussakin kaupungissa. Kuntien tilikarttojen rakenteen on määrittänyt historia ja kirjanpidossa käytettävä ohjelmisto ja

³⁹ <http://frictionlessdata.io/>

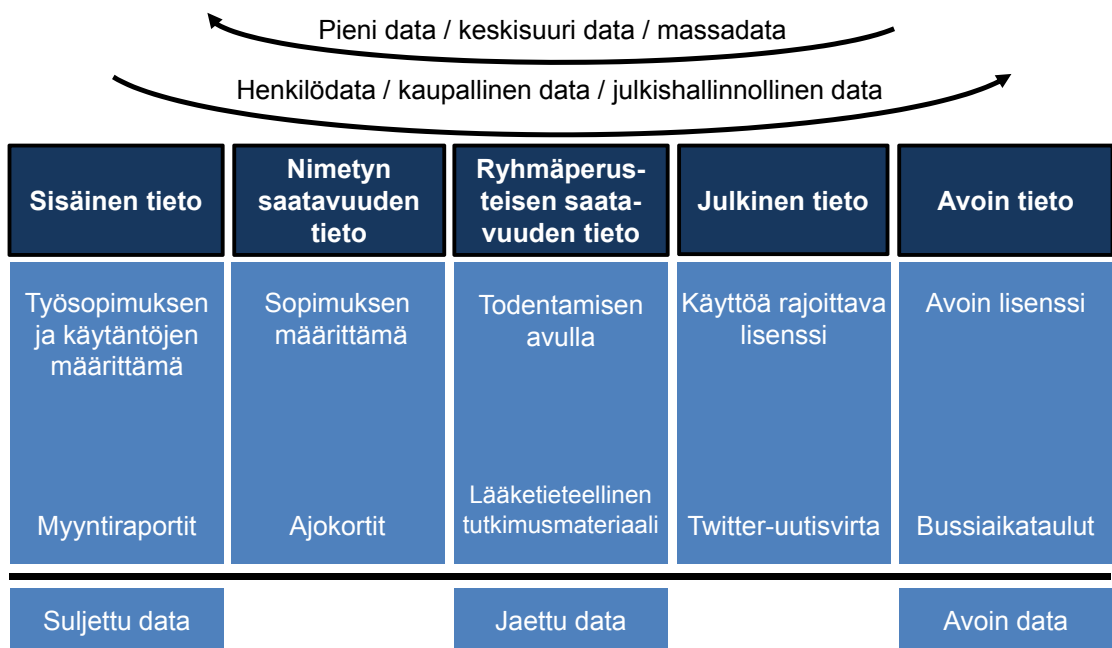
⁴⁰ <https://www.force11.org/group/fairgroup/fairprinciples>

hankintakulttuuri, jolloin niiden yksiselitteinen ymmärtäminen tai vertailtavuus keskenään on suurelta osin mahdotonta. Yhteisen standardoidun tietomallin ja hankintasopimuksien avulla kuntien kustannustehokkuutta voisi vertailla.

4.6.4 Dataspektri kuvaa tiedon avoimuuden koko skaalaa

Usein keskustellaan siitä, voidaanko jotain dataa avata ja julkaista, kun voisi olla hyödyllisempää pohtia, voidaanko yhtenä vaihtoehtona käyttöä lisätä ”avaamalla dataa vain osittain”. Open Data Institute (ODI) havainnollistaa datan avoimuuden tasoja dataspektrillä⁴¹ (data spectrum).

Kuvio 4.3 Dataspektri



Lähde: <https://theodi.org/data-spectrum>

Dataspektrissä datan avoimuutta kuvataan viidellä tasolla, suljetusta avoimeen:

- **Sisäinen tieto (Internal access)** on tyypillisesti salaista tai rajattua tietoa oman tai organisaation sisäiseen käyttöön – kuten esimerkiksi myyntiraportit, työsopimukset tai henkilökohtainen tieto kuten terveystieto.
- **Nimetyin saatavuuden tieto (Named access)** on tietoa, jonka saa erillistä sopimusta vastaan. Tämä viittaa esimerkiksi tietoon, jota kaupalliset toimijat saavat julkishallinnon organisaatiolta (kuten henkilötiedot Väestörekisterikeskukselta), ja jota hyödynnetään tai jatkojalostetaan kaupallisiksi tuotteiksi.
- **Ryhmäperusteisen saatavuuden tieto (Group-based access)** on tietoa, joka on saatavilla esimerkiksi todentamisen perusteella. Julkishallinnon organisaatioiden välinen tiedonvaihto kuuluu tähän ryhmään.

⁴¹ <https://theodi.org/data-spectrum>

- **Julkinen tieto (Public access)** on tietoa joka on julkisesti saatavilla mutta ei ole täysin avointa, esimerkiksi koska sen uudelleenkäyttöä rajoitetaan lisenssiehdoin, tai tieto ei täytä kaikkia avoimen määritelmän kriteerejä. Esimerkiksi sosiaalisen median palveluiden sisällöt (Twitter, LinkedIn, Facebook, ym.) eivät ole täysin avoimia. Myös joissain julkishallinnon tietoaaineistoissa on saatettu jättää mainitsematta lisenssi, tai lisenssi ei salli datan kaupallista käyttöä, jolloin ei voida sanoa, että kyseessä on ”avoin tieto”.
- **Avoin tieto (Anyone)** on dataa tai tietoa, joka täyttää avoimen tiedon määritelmän, ja on siten vapaasti käytettävissä, myös kaupallisesti. Suurin osa haastatelluista organisaatioista julkaisee jotain avointa dataa, tyypillisesti JHS 189:n mukaisella Creative Commons 4.0 -lisenssillä.

Dataspektri voi olla hyödyllinen työkalu arvioitaessa eri tietoaaineistojen avaamista. Yksi olennainen kysymys hyödyntämisen lisäämiseksi onkin, että ”mitä voidaan tehdä, että siirryttäisiin dataspektrillä *kohti* oikea laitaa” – ei välttämättä oikeaan laitaan. Tiedon laajempi hyödyntäjäkunta lisää käyttöpotentiaalia.

4.6.5 Henkilötiedon hyödyntäminen

Avoimen datan ei tulisi sisältää henkilötietoja, liikesalaisuuksia tai muuta luottamuksellista tietoa. Merkittävä osa julkisista tietovarannoista kuitenkin sisältää henkilötietoja, jolloin niitä ei voi suoraan avata avoimeksi dataksi. Yhtenä julkisen sektorin datamassojen käsittelyn suurimpana riskinä pidetäänkin henkilötietojen suojan murtumista (Antikainen et al., 2016, s. 34).

Yleisesti ottaen tiedon laajempi hyödyntäjäkunta lisää käyttöpotentiaalia. Henkilötietoja sisältävien aineistojen hyödyntäjäkuntaa voidaan laajentaa ilman, että yksityisydensuojasta tingitään. Alunperin salassapidettäviä tietoja sisältäviä aineistoja voidaan anonymisointimenetelmillä muuttaa siten, että ne ovat avattavissa, kuten on tehty esimerkiksi ajoneuvorekisterin tapauksessa. Toinen mahdollisuus on, että ihmisille itselleen annetaan mahdollisuus päättää häntä koskevan henkilötiedon käytöstä omadata- eli MyData-periaatteiden mukaisesti. Kuviossa 4.4 on havainnollistettu henkilötiedon ja avoimen datan suhdetta tiedon hyödynnettävyyteen.

Anonymisoinnissa dataa käsitellään siten, että henkilötiedot, liikesalaisuudet ja muu luottamuksellinen tieto häivytetään pois ennen datan julkaisua. Datan tarkkuutta voidaan esimerkiksi laskea, dataan voidaan lisätä satunnaisuutta ja tiettyjä tietoja (kenttiä tai sarakkeita) voidaan poistaa siitä (Green et al., 2017).

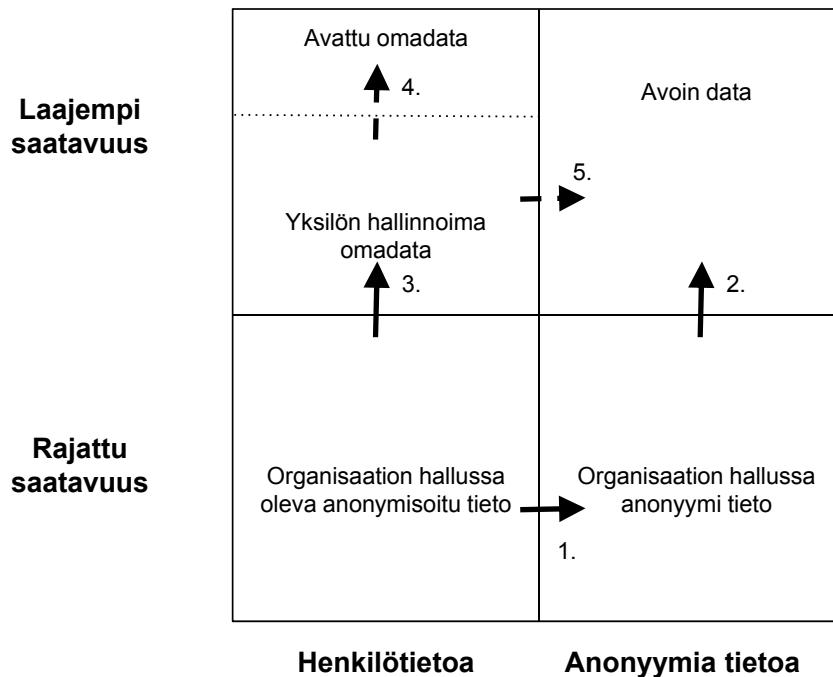
Tieto on anonymiä, jos tunnusomaiset piirteet koskevat samanlaisina useampaa henkilöä ja siitä ei voi millään kohtuullisesti toteutettavissa olevilla keinoilla tunnistaa yksittäisiä henkilöitä. Anonymisoinnin tulee olla peruuttamaton, jotta voidaan puhua anonymistä aineistosta. Aineisto on henkilötietolain mukaisesti tunnistettavissa niin kauan kun koodiavain henkilötietoihin on olemassa, vaikka sitä säilytettäisiin aineistosta erillään.

Anonymistä aineistosta ei voi tunnistaa yksittäisiä henkilöitä epäsuorien tunnistajien avulla. Esimerkiksi harvaan asutulla alueella tieto rakennuksesta ei ole anonymiä tietoa, jos sen perusteella voidaan yksilöllisesti tunnistaa rakennuksen asukas. Myöskään henkilön tunnistaminen yhdistämällä aineistoon muualta saatavia tietoja ei tulisi olla mahdollista. Yhdistämällä anonymiksi tarkoitettu avoin tieto johonkin muuhun aineistoon on datan anonymiteetti onnistuttu joissain tapauksissa osittain tai kokonaan purkamaan (Green et al., 2017). Anonymisoinnin keskeinen haaste onkin, että erilaisia tietoja yhdistellään yhä monipuolisemmin toisiinsa,

tietoa on saatavilla yhä enemmän ja aineistojen yhdistelytapoja ja teknologista kehitystä (esimerkiksi tekoäly & algoritmit) ei voida täysin ennakoida etukäteen.

Anonymisointi väistämättä heikentää datan hyödynnettävyyttä joihinkin käyttötarkoituksiin, esimerkiksi personoituja dataan pohjautuvia palveluja voi tuottaa vain silloin, kun datassa on tallella yksilöivät tunnisteet. Anonymisointikeinoja valittaessa pyritään kuitenkin siihen, että esimerkiksi datan tilastolliset ominaisuudet eivät muutu merkittävästi ja että datan informaatioarvo pysyy mahdollisimman korkeana. Datan anonymisoinnin asiantuntemusta on Suomessa esimerkiksi Trafissa ja Tilastokeskuksessa. Avoimen datan anonymisointiosaaminen on esimerkki kansainvälisestikin tärkeästä uudesta osaamisalueesta. Olisi varmasti löydettävissä hyviä käytäntöjä, jotka voitaisiin toteuttaa Euroopan laajuisina, mm. yhtenäistyvän

Kuvio 4.4 Henkilötietoa voidaan käsitellä niin, että siitä tulee avointa dataa joko henkilön omalla suostumuksella tai anonymisoimalla



Kuvio on suuntaa antava hahmotelma asiasta.

- Anonymisointi:** tietovarannon hallinnoija voi anonymisoida datan, jolloin se ei enää sisällä henkilö-tietoja
- Anonyymien datan avaaminen:** tietovarannon hallinnoija voi anonymisoida datan ja julkaista sen avoimena datana (esim. Trafista avoimet ajoneuvo- ja venerekisterit)^a
- Henkilölle pääsy omaan dataan:** tietovarannon hallinnoija antaa ihmisille pääsyn heitä itseään koskevaan dataan ja nämä voivat halutessaan uudelleenkäyttää omadataa muissa palveluissa, kuten henkilökohtaisissa avustajasovelluksissa.
- Henkilö avaa omaa dataansa:** ihminen voi joissain tapauksissa luovuttaa omaa henkilötietoaan avoimeksi dataksi vaikkapa vertaistiedoksi kuntoilusovellukseen
- Henkilö avaa anonymisoidun omadatan:** henkilö voi luovuttaa oman datansa anonymisointuna esimerkiksi erilaista tutkimusta varten.

^a https://www.trafi.fi/tietoa_trafista/ajankohtaista/4541/trafin_avoin_data_kattaa nyt_kaikki_liikennemuodot

tietosuojalainsäädännön myötä. Kuviossa 4.5 on kuvattu Trafín menetelmä luoda käyttörajoitetusta tiedosta avointa dataa.

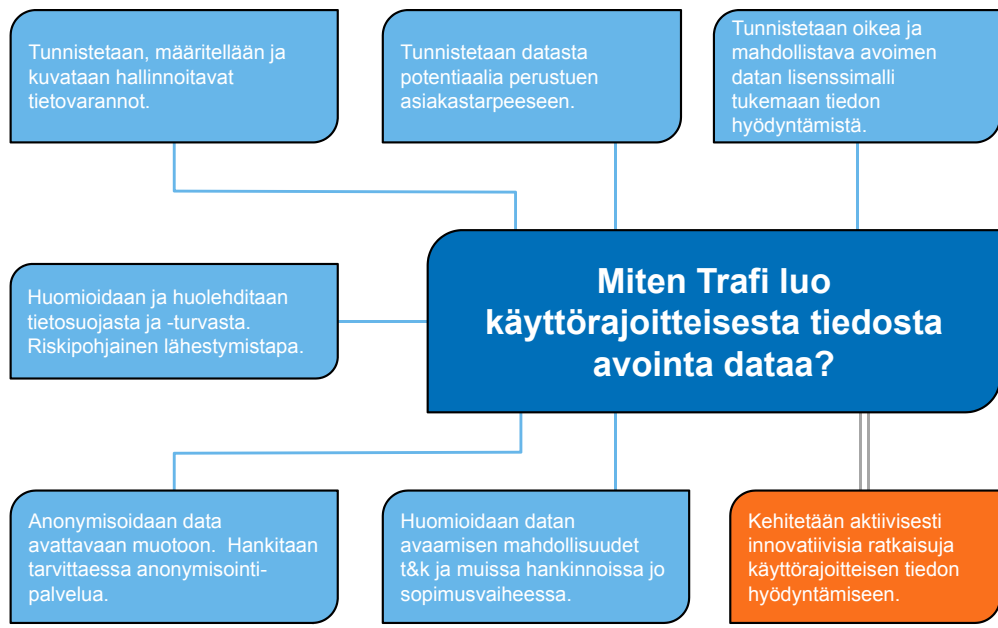
Anonymisointi jakaa mielipiteitä. Jotkut tutkijat väittävät, että melkein kaiken datan voi anonymisoida, kun taas toiset ovat sitä mieltä, ettei täysin luotettavaa datan anonymisointia ole, vaan mitä enemmän dataa julkaistaan, sitä helpompaa on selvittää identiteettiä datasta yhdistelemällä eri tietoaaineistoja keskenään. Lopulta kyse on yhteiskunnan arvoista ja riskeistä. Millaisia riskejä voidaan ottaa yksityisyydensuojan menettämisen suhteen, jos kannustimena on esimerkiksi parempi lääketieteellinen hoito? (Economist, 2015).

Myös haastatteluissa nousi esille huoli yksityisyyden suojasta ja avoimen datan julkaisun turvallisuudesta pitkällä aikavälillä, jos eri aineistojen yhdistelyyn pohjautuvat niin sanotut de-anonymisointimenetelmät kehittyvät. Toisaalta haastatteluissa nousi avoimesta datasta vahvasti esille näkemys, että ”suurimmat innovaatiot syntyvät tietojen yhdistämisestä”.

Omadata eli MyData

Henkilötiedon kohteella on nykyisin lähinnä passiivinen rooli omien henkilötietojensa hallinnassa ja hyödyntämisessä. Yleensä henkilö ei saa itse päättää laajasti omien henkilötietojensa käytöstä, eikä tietoja ole helppoa saada omaan käyttöön sähköisessä ja rakenteellisessa muodossa.

Kuvio 4.5 Case Trafi: Käyttörajoitteisesta datasta avointa dataa



Trafín tavoitteena on tehdä käyttörajoitteisesta tiedosta avointa dataa (Parkkisenniemi, 2017). Käyttörajoitteisen tiedon avaamista lähestytään lisenssimallin valinnan, riskipohjaisen lähestystavan ja anonymisoinnin avulla. Datán avausprosessia priorisoidaan mallissa asiakastarpeen pohjalta. Asiakastarpeeseen perustuva priorisointi onnistuu, mikäli asiakkaat tietävät millaisia tietovarantoja organisaatio hallinnoi. Sen takia mallin osana on myös hallinnoitavien tietovarantojen tunnistaminen, määrittely ja kuvaaminen.

Omadata (MyData) on ihmiskeskeinen henkilötiedon hallinnan ja käsittelyn periaate, jonka mukaan ihmisille tarjotaan mahdollisuus hallinnoida, hyödyntää ja päättää heistä kerätyn henkilötiedon (kuten esim. ostostiedot, liikkumistiedot, taloustiedot, terveystiedot ja eri verkopalveluihin kertyvä data) luovuttamisesta eteenpäin (Poikola et al., 2014). Tässä mallissa omadata mielletään yhdeksi henkilödatan osajoukoksi riippuen siitä, onko yksilöllä itsellään mahdollisuutta hyödyntää ja hallinnoida sitä. Näin ollen kaikki omadata on henkilötietoa, mutta kaikki henkilötieto ei ole omadataa.

Oman datan hallinnointia voi tapahtua useammalla eri tavalla. Ihminen voi esimerkiksi luvittaa omien osoitetietojen jakelua eri toimijoille kuten pankki, posti, matkapuhelinoperaattori, kaupan toimijat, suoramarkkinointi jne. Ajoneuvorekisterin tietoja voisi luvittaa esimerkiksi vakuutus- ja katsastusliikeille tarjouksia varten, erilaisiin tutkimukseen tai korjaamoille. Kiinteistörekisterin tietoja voisi luvittaa esimerkiksi naapureille, yhteisiin hankkeisiin kuten esimerkiksi tien rakentaminen tai luonnonsuojeluun. MyData-mallissa henkilö käyttötapauskohtaisesti antaa suostumuksensa omien tietojensa hyödyntämiseksi luotetuille tahoille tai jopa kaikille avoimeksi.

Tällainen oman datan luvittaminen laajentaa henkilötiedon käyttötapauksia ja siten voi lisätä datan vaikuttavuutta. Henkilön itsensä tekemän luvittamisen ohella muut nykyiset lailliset henkilötiedon käsittelyn perusteet säilyvät. Haastattelussa ja työpajoissa omadataa pidettiin tärkeänä osana erilaisten tietovarantojen kokonaisvaltaista hyödyntämistä (avoin data, masadata ja omadata). EU tietosuojasetuksen myötä kiinnitetään entistä enemmän huomiota henkilötietoasioihin.

5. TOIMENPITEITÄ AVOIMEN DATAN VAIKUTTAVUUDEN LISÄÄMISEKSI

Tässä luvussa esitetään ehdotuksia siitä, millaisia politiikkatoimia, toimintatapoja ja rekisterien rakenteellisia muutoksia tarvitaan julkisen hallinnon tietovarantojen hyödyntämisen tehostamiseksi. Ehdotukset perustuvat kyselytutkimuksiin, haastatteluihin, työpajoihin ja muuhun hankkeessa tehtyyn selvitystyöhön. Kussakin suosituksessa kuvataan itse suositus, tavoitteet sekä toteuttaja.

Suosituksilla tähdätään seuraaviin vaikuttavuutta lisääviin yleisiin tavoitteisiin:

1. Hallittu ja selkeä datan hyödyntämisen johtaminen ja ohjaus
2. Lisäännytyn datatalouden osaaminen ja yhteistyö eri toimijoiden kesken
3. Selkeät säädökset ja niitä tukevat käytännöt
4. Laadukas tietoinfrastruktuuri

Suositus 1: Yhtenäistetään tietojohdamisen työkaluja

Kuvaus: Tietointensiivisen toiminnan johtaminen edellyttää kokonaisvaltaista, vertailukelpoista ja oikeaa tietoa organisaation tiedonhallinnasta. Selkeä, jäsennetty ja raportoitu tiedonhallinta on perusedellytys toiminnan kehittämiseksi ja luotettavalle tutkimukselle.

Sekä tulossopimuksilla että tietotilinpäätöksillä tulisi olla yhtenäinen standardoitu asiakirjakenne ja ne tulisi julkaista rakenteellisessa ja koneluettavassa muodossa yhdessä paikassa. Niiden standardoinnissa tulisi pyrkiä siihen, että eri hallinnonalojen ja virastojen asiakirjat ja niiden osat olisivat mahdollisimman helposti vertailtavissa keskenään, vastaavasti kuin taloudellisessa kirjanpidossa on helppo vertailla eri organisaatioiden tilinpäätöstä ja tasetta. Tietotilinpäätöksen tulisi viestiä selkeästi organisaation koko tiedonhallinnan kokonaiskuva, sisältäen kuvauksia ja mittareita mm. tietojärjestelmistä, tietovarannoista, tietovirroista, kansallisen palveluväylän liittymistä, rekisteriselosteista, ohjelmointirajapinnoista, tietohallintoon liittyvistä projekteista, avoimen datan tarjonnasta ja tietojenkäsittelyyn vaikuttavasta keskeisestä lainsäädännöstä.

Taulukko 5.1 Avoimen datan mittaaminen osana tulossopimusta (tulevaisuus) ja tietotilinpäätöstä (jo tapahtunut) – esimerkki mahdollisista mittareista

Mitattava asia	Esimerkki mittarista	Lisätietoja
1. Avoin data <i>(avattu data)</i> <i>(avoimen datan potentiaali)</i>	Kuinka monta prosenttia kokonaispotentiaalista on avattuna ja löydettävissä?	Kartoitetaan organisaation datakokonaisuus, joka voisi olla avointa. Lukua vasten laskeetaan osuus, joka on jo avointa.
2. Avoimen datan hyödyntäjien asiakastytyväisyys <i>Likert-asteikko</i>	Arvosana (esim. 1–5) jonka avoimen datan hyödyntäjät antavat datan avaajalle.	Datan avaajien ja hyödyntäjien välinen vuorovaikutus on tärkeää, mittari ohjaa kuuntelemaan datan palautetta ja oppimaan siitä.
3. Omadata (tulevaisuus) <i>(luvittamisen piirissä oleva omadata)</i> <i>(organisaatiolla oleva omadata)</i>	Missä määrin asiakkaille on annettu mahdollisuus hallinoida organisaation rekistereissä olevaa itseään koskevaa tietoa?	Omadata on henkilötietojen se osuus, jota käyttäjä voi itse hallita.

Tavoite: Kokonaisvaltainen ja järjestelmällinen toiminta, tehostettu tietotyö, kokonaisuuden hahmottaminen, virastojen ja laitosten vertailtavuus.

Toteutus: Suositus voidaan toteuttaa kaikissa ministeriöissä, virastoissa ja laitoksissa. Avoin datan kustannustehokkaan hyödyntämisen kannalta hyviä jatkokehitettäviä toiminnan tilaa kuvaavia mittareita on esitetty taulukossa 5.1.

Tieto- ja avoimuuspolitiikan ja datan hyödyntämisen suunnitelman julkaiseminen avoimesti

Kuvaus: Siinä missä tulossopimus ja tietotilinpäätös ovat erinomaisia johtamisen välineitä, olisi hyödyllistä laatia selkeä ja ulospäin viestitty tahtotila ja suunnitelma eri organisaatioiden tieto- ja avoimuuspolitiikasta ja datan hyödyntämisestä. Tietovarantojen avaamisen lisäksi datan hyödyntämissuunnitelman olisi hyvä kattaa myös massadatan, omadatan ja käyttörajoitettujen datan hyödyntäminen. Tavoitteeksi asetetaan *lähtökohtaisesti digitaalisesti avoin (open by default) -periaate* ja luodaan yhtenäiset politiikat tätä tukemaan. Kaiken datan tulisi olla avointa, ellei ole erikseen perusteltua syytä rajoittaa tietoon pääsyä esimerkiksi lainsäädännöllisen syyn johdosta. Panostetaan Open Government Partnership -ohjelmaan, Open Data Charter -julistukseen ja muihin kansainvälisiin aloitteisiin, joilla tehdään Suomen sitoutuminen selväksi ja toiminta kansainvälisesti vertailukelpoiseksi.

Tavoite: Järjestelmällisyys datan hyödyntämiseen, ennakoitavuus, datan hyödyntämisen laajentaminen, uusi liiketoiminta, tahtotilan jalkauttaminen, nykyistä yhtenäisemmät käytännöt.

Toteutus: Suositus voidaan toteuttaa kaikissa ministeriöissä, virastoissa ja laitoksissa. Olenaisilta osin mukaan osaksi tulossopimusta.

Suositus 3: Järjestelmällinen tietovarantojen kartoitus ja niiden hyödyntämispotentiaalin arviointi

Kuvaus: Arvioidaan ministeriöiden, virastojen ja laitosten potentiaalisia korkean lisäarvon data- tai tietoaineistoja. Tämä voidaan tehdä esimerkiksi osana tietosuoja-asetukseen varautumiseen liittyviä tietovarantokartoituksia ja KaPa-projekteja. Samalla voidaan kartoittaa, ovatko henkilötietoa sisältävät tietovarannot julkaistavissa osin avoimeksi, kuten Maanmittauslaitos teki kiinteistörekisterin osalta⁴², tai onko data muuten hyödynnettävissä, esimerkiksi anonymisoimalla. Hyödynnetään EU:n suosittelemia käytäntöjä, kehittäjäyhteisöjä, datanhyödyntäjiä ja tietopyyntötilastoja uusien tietoaineistojen avaamisen priorisointiin. Priorisoinnin perusteella päivitetään datan avaus- ja hyödyntämissuunnitelmat, joiden tulisi olla saatavilla keskitetysti esimerkiksi Avoindata.fi:ssä.

Tavoite: Järjestelmällisyys datan hyödyntämiseen, priorisointi, datan hyödyntämisen laajentaminen, uusi liiketoiminta korkean lisäarvon tietoaineistojen avaamisesta.

Toteutus: Suositus voidaan toteuttaa kaikissa ministeriöissä, virastoissa ja laitoksissa. Olenaisilta osin mukaan osaksi tulossopimusta.

⁴² <http://www.maanmittauslaitos.fi/ajankohtaista/kiinteistorekisterikartta-avoimeksi-dataksi>

Suositus 4: Panostetaan ketterään, avoimeen ja kokeilevaan toimintakulttuuriin

Kuvaus: Lisätään vuorovaikutusta julkishallinnon, yritysten, kolmannen sektorin ja muiden datan hyödyntäjien kanssa. Ketterällä kokeilukulttuurilla tavoitellaan mm. liiketoiminnan pilotointia, avaamistapojen pilotointia, iteratiivista toimintamallia ja käyttäjälähtöisyyttä. Samalla luodaan ja varmistetaan onnistuneille kokeiluille jatkopolkuja, joissa edistäjinä voivat olla muutkin tahot kuin kokeilun aloittaja. Yhtenä esimerkkinä tästä on Trafín Liikennelabran toiminta.

Julkaistaan dataa tai rajapintoja kokeiluluontoisesti, kattavasti ja ripeästi, raa'assakin muodossa ja tarjotaan palautemekanismeja, joiden avulla parannetaan datan tai rajapintojen laatua. Esimerkki toimintamallista: i) avataan dataa kokeiluluontoisesti, ii) osallistutaan hackathoniin tai vastaavaan tapahtumaan, iii) tehdään nopea kokeilu iv) korjataan laatua, mikä johtaa datan tai rajapinnan varsinaiseen julkaisuun.

Lisätään yhteistoimintaa kehittäjäyhteisön kanssa tapahtumissa ja verkossa, jolloin päästään paremmin kiinni hyödyntäjien tarpeisiin. Sisällytetään hackathonit, kilpailut, muut tapahtumat johdonmukaiseksi toiminnaksi. Hyödynnetään verkossa avoimia sähköisiä vuorovaikutteisia foorumeita järjestelmällisesti ja suunnitelmallisesti, joilla lisätään näkyvyyttä ja suoraa vuorovaikutusta.

Tavoite: Datan laajempi hyödyntäminen, tietovarantojen markkinointi, ketterä liiketoiminnan pilointi, nopea oppiminen suoran vuorovaikutuksen kautta, parhaiden käytäntöjen jakaminen, uusi liiketoiminta, käyttäjälähtöisyys.

Toteutus: Suositus voidaan toteuttaa kaikissa ministeriöissä, virastoissa ja laitoksissa. Vastuutetaan kehittäjäyhteistöimintaa yksittäisten organisaatioiden lisäksi esim. Avoindata.fi:lle tai toimialat ylittävälle yhteistyöverkostoille kuten Lynet ja Soterko. Kokeilutoimintaan otetaan mukaan mm. Muutoksentekijät -verkosto ja Kokeilun Paikka⁴³ sekä muut kehittäjäyhteisöt ja verkostot. Kuntien rooli on myös tärkeä.

Suositus 5: Datatalouden ymmärtäminen ja datan osaamisen vahvistaminen

Datatalouden mekanismien ymmärtämiseksi tarvittaisiin tutkimustietoa yritysten datan käytöstä ja taloudellisista vaikutuksista. Tilastoaineistoja tulisi kerätä seuraavista seikoista:

- Mitä dataa yritykset käyttävät?
- Mihin tarkoituksiin dataa käytetään? Esim. innovaatiotoiminta: uusien tavaroiden tai palveluiden kehittäminen tai tuotteiden parantaminen, prosessi-, organisaatio- ja markkinainnovaatiot, logistiikkaan liittyvät innovaatiot; tuotantoprosessin hallinta, markkinointi.
- Ostaako tai myykö yritys (big) dataa?
- Keinoälyn rooli datan hyödyntämisessä

Erilaisiin kyselytutkimuksiin tulisi lisätä sellaisia keskeisiä kysymyksiä, jotka lisäävät datatalouden ymmärrystä. Tällainen kysymys on esimerkiksi se, miten avoin data edesauttaa

⁴³ www.kokeilunpaikka.fi/

avoimen kehittämisen yhteisöjä, alustoja tai ekosysteemejä ja siten uudentyypisiä liiketoimintamalleja ja ansaintalogiikkaa (esim. liikkumispalveluja tarjoavat alustat). Datatalouden merkityksestä kilpailulle (mm. toimialarajojen hämärtyminen) laajemmin tarvitaan lisää tietoa. Myös kuluttajien yksityisyydensuojaa koskevat kysymykset ovat tärkeitä.

Avoimen datan ja laajemmin datatalouden parhaita käytäntöjä tulisi levittää mm. verkkokursseilla tai laajemmin osaamisohjelmalla. Erityishuomioita tulisi kiinnittää datan luku-, tulkinta- ja käsittelytaitoihin, jotka ovat tulevaisuudessa lukutaitoon rinnastettavia taitoja. Lisätään kansainvälistä yhteistyötä ja oppimista, mm. Open Data Charter ja Open Government Partnership -ohjelmilla.

Datan tietosuoja- ja anonymisointiosaamiseen panostaminen on erityisen tärkeää, sillä henkilötietoa sisältävillä tietovarannoilla arvioidaan olevan korkea lisäarvopotentiali. Myös henkilötietoja sisältävien tietovarantojen osittainen avaaminen tai laajempi käyttöluvittaminen edistää niiden laajempaa käyttöä.

Ohjelmistojen kehitys- ja hankintaosaamista tulee kehittää. Erityisesti räätälöityjen ohjelmistojen hankinnassa tulisi hyödyntää avointa lähdekoodia, joka tukee datan hyödyntämistä. Esimerkiksi Ilmatieteen laitos ja Liikennevirasto tilaavat lähtökohtaisesti kaikki räätälöidyt ohjelmistot avoimina. Huomioidaan avoin data ja avoimet rajapinnat tietojärjestelmähankinnoissa.

Tavoitteet: Lisääntynyt julkishallinnon dataosaaminen, henkilötietoa sisältävien korkean lisäarvopotentialin omaavien tietovarantojen laajempi hyödyntäminen, kotimainen ja kansainvälinen parhaiden käytäntöjen levittäminen.

Toteutus: Hyödynnetään verkostoja kuten valtiovarainministeriön osaamisverkosto, Big Data Forum, MyData-allianssi, jne. Koulutustoiminta ja osaamisen kehittäminen voisi olla esimerkiksi Väestörekisterikeskuksen tai mahdollisesti D9:n, JulkICT:n tai jonkin muun poikkihallinnollisen toimijan vastuulla. Lisäksi tarvitaan yhteistyötä yliopistojen ja oppilaitosten kanssa. Datat hyödyntämistä koskevia kysymyksiä tulisi lisätä Tilastokeskuksen säännöllisesti toteutamiin kyselytutkimuksiin linjassa yllä esitetyn listauksen ja luvun 3 kanssa.

Suositus 6: Viranomaistoimija huolehtimaan avoimesta tiedosta, edistämään data hyödyntämistä ja yhtenäistämään juridisia käytäntöjä

Kuvaus: Datat tehokas hyödyntäminen tarvitsee keskitetyn toimielimen, joka linjaa juridisia käytäntöjä yli hallintorajojen. Tämä toimielin (esimerkiksi tietovaltuutettu, avoimen tiedon valtuutettu tai datavaltuutettu) vastaisi julkisuuslain (621/1999), avoimen tiedon ja omadatan edistämisestä ja toimisi eräänlaisena tietosuojavaltuutetun vastinparina esimerkiksi tiedon avaamiseen liittyvien käytäntöjen ja suositusten muodostamisessa. Toimielimen tehtävänä olisi myös antaa ohjeistuksia tiedon ja tietovarantojen anonymisoinnin osalta esimerkiksi massaluovutusta varten, jotta yksityisyydensuoja ja tiedon avoimuus turvataan. Toimielimen tehtäviin kuuluisi myös tulkinta-ohjeiden tuottaminen lisenssiasioiden, maksuperustelain ja muiden tietovarantoihin ja avoimeen dataan liittyvien oikeudellisten asioiden tulkinnasta. Vaihtoehtoisesti tässä suosituksessa kuvatut tehtävät voisi lisätä esimerkiksi tietosuojavaltuutetun toimiston tehtäviin.

Tavoite: Datan hyödyntämisen tehokkuus, yhtenäiset käytännöt, selkeä sääntely, esteiden purku.

Toteutus: Uusi viranomainen, uusi tehtävä tietosuojavaltuutetulle tai uusi tehtävä jollakin ICT- tai digitalisaatio-organisaatiolle.

Suositus 7: Avoindata.fi-dataportaalin jatkokehittäminen

Kuvaus: Julkishallinnon organisaatioiden avoimen datan löydettävyyden edistämiseksi tietoa avatusta datasta tulisi koota kattavasti ja linkitetysti Avoindata.fi:hin ja joihinkin muihin palveluihin, joista avattu data olisi siirrettävissä mm. EU:n avoimen datan portaaliin. Avoimen datan kansainvälisen löydettävyyden kannalta aineiston metatiedot tulisi antaa myös englannin kielellä, mikä parantaisi Suomen näkyvyyttä kansainvälisissä vertailuissa, lisäisi Euroopan avoimen datan portaalissa olevien aineistojen käytettävyyttä sekä palvelisi kansainvälisiä avoimen datan kehittäjiä ja muuta yleisöä. Avoimen datan käytön kasvaessa ja aineistojen saatavuuden laajentuessa olisi hyvä kehittää Avoindata.fi-palvelun käytettävyyttä, palvelutasoa ja tukipalveluita sekä markkinoida sitä julkisten organisaatioiden lisäksi myös yritysten ja kolmannen sektorin julkisen datan avaamiseen.

Tavoite: Toimiva tietoinfrastruktuuri, datan löydettävyys paranee, datan käyttö Suomessa ja kansainvälisesti lisääntyy, määritelty palvelutaso ja -lupaus tukevat dataan perustuvaa liiketoimintaa, Suomi sijoittuu kymmenen kärkimaan joukkoon kansainvälisissä vertailuissa.

Toteutus: Koskee ensisijaisesti Väestörekisterikeskusta ja sen Avoindata.fi-palvelua.

Suositus 8: Tietovarantojen käytettävyyden ja avoimen datan laadun jatkuva parantaminen

Kuvaus: Datan lisenssi, versiointi, tiedostoformaatti, metadata, datan sisältämät tunnisteet, käytetyt sanastot, saatavuus ja monet muut asiat vaikuttavat olennaisesti datan laatuun, lisäarvoon ja hyödynnettävyyteen. Jatketaan metadata- ja käsitelyä ja semanttisen yhteensopivuuden hankkeita mahdollisimman avoimesti. Kunnat tuottavat merkittävän osan perusrekistereiden sisältämisestä tiedoista ja kunnat myös vastaavasti hyödyntävät perusrekistereitä huomattavan paljon, joten tietomallien, metadatan, käsitteistön, tietovirtojen ja ohjelmointirajapintojen yhteensopivuuteen ja harmonisointiin kuntasektorilla tulee panostaa. Tutkimusdatan laatua, käyttöoikeuksia (lisenssit) metadataa ja löydettävyyttä tulee parantaa soveltaen mm. FAIR-periaatteita. Myös yleisesti niin maksullisten, luvanvaraisten kuin avoimien tietovarantojen kattavat metatiedot ja läpinäkyvyys muun muassa tietotuotteiden hinnoittelun ja tiedotomuuksien suhteen parantaisivat tietovarantojen hyödynnettävyyttä. Nämä tiedot voitaisiin julkaista keskitetysti esimerkiksi Avoindata.fi:ssä.

Tavoite: Yhteensopivuus, löydettävyys, uusi liiketoiminta, tehostaminen.

Toteutus: Suositus voidaan toteuttaa kaikissa ministeriöissä, virastoissa ja laitoksissa. Keskitetty tietovarantojen tai tietotuotteiden katalogi maksuperusteineen voitaisiin toteuttaa esimerkiksi Avoindata.fi:ssä.

Suositus 9: Avoimet ohjelmointirajapinnat ja keskitetty rajapintojen hallinta sujuvoittamaan yhteensopivuutta ja edistämään kilpailua

Kuvaus: Ohjelmointirajapintojen dokumentaatio, hallintamalli ja palvelulupaus (SLA) pitäisi ohjeistaa nykyistä paremmin. Säädöksissä tai hankintaohjeistuksissa tulee huomioida tai määritellä ohjelmointirajapinnan avoimuutta koskevat minimivaatimukset, jotta voidaan välttää tehottomuutta ja toimittajariippuvuutta. Riippuvuus ohjelmistotoimittajista ja tietovirtojen lukitseminen (suljetuilla) ohjelmointirajapinnoilla haittaavat hallituksen ”tietoja kysytään vain kerran” -periaatetta ja yleisemmin monipuolisen palvelutarjonnan kehittymistä. Ohjelmointirajapinnoille tulisi perustaa kansallinen keskitetty ”digitaalinen puhelinluettelo” eli katalogipalvelu, jossa tulisi olla avoimesti saatavilla kunkin ohjelmointirajapinnan hallintamalli, palvelulupaus, rajapintakuvaukset ja tietomallit. Keskitetty rajapintojen hallinta (API management), jonka piiriin kuuluisi Kansallisen palveluväylän liityntäkatalogi (kansallisen palveluväylän liitynnät), avoimet ohjelmointirajapinnat ja julkishallinnon ohjelmointirajapinnat, tarjoaisi keskitetysti tietoa myös tietovirroista ja tietovarantojen käytöstä.

Tavoite: Datan yhtenäistäminen, laadun parantaminen, työmäärän väheneminen, datan löydettävyyden parantaminen, datan hyödyntämisen mittaaminen, kilpailun lisääntyminen.

Toteutus: Suositus voidaan toteuttaa kaikissa ministeriöissä, virastoissa ja laitoksissa. Keskitetty rajapintakatalogi voisi olla erillinen palvelu, joka täydentäisi Avoindata.fi-palvelua.

Suositus 10: Mahdollistetaan vapaaehtoinen avoimen datan hyödyntäjien rekisteröityminen

Kuvaus: Datan hyödyntäjät ovat riippuvaisia datan avaajista. Datan hyödyntäjät tarvitsevat tietoa muuttuneista ohjelmointirajapinnoista tai tietoaineistojen päivityksistä taatakseen kehittämiensä palveluiden laadun ja jatkuvuuden. Toisaalta tiedon julkaisijat eivät aina saa tietoa datan käyttäjistä, vaikka laajasti käytössä oleva JHS 189:n mukainen CC BY 4.0 -lisenssi velvoittaa datan hyödyntäjiä nimeämään datan lähteen. Vapaaehtoisella rekisteröitymisellä lisätään vuorovaikutusta, datan hyödyntäjä voi luottaa saavansa tietoa muuttuneesta tietolähteestä ja datan julkaisija saa tietoa dataa hyödyntävästä sovelluksesta. Rekisteröityminen voidaan tehdä esimerkiksi verkkolomakkeella automaattisesti ja avoin data siis olisi edelleen saatavissa myös ilman rekisteröitymistä.

Tavoite: Datan julkaisijoiden ja hyödyntäjien lisääntynyt yhteistyö, vaikuttavuuden arvioinnin parantaminen, luottamus datan julkaisijaan.

Toteutus: Suositus voidaan toteuttaa kaikissa ministeriöissä, virastoissa ja laitoksissa. Toiminnallisuuden voi liittää avoimen datan palveluihin.

6. PÄÄTELMÄT

Avoimen datan taloudellisen vaikuttavuuden tutkimisen suurimpana haasteena on tällä hetkellä systemaattisen tiedonkeruun puute avoimen datan hyödyntämisestä. Tietoja avoimen datan hyödyntäjistä voitaisiin hankkia avoimen datan käyttäjäportaaleista. Merkittävin este käyttäjien tunnistamiselle kyselytutkimuksessa mukana olleiden rekisterinpitäjien joukossa on tiukka avoimen datan käytön anonyymisyyden periaatteen noudattaminen. Käyttäjänäkökulmasta avoimen datan käyttötietojen kerääminen pelkästään tutkimustyössä käytettäviin tietokantoihin ei tunnu kovinkaan radikaalilta ratkaisulta maailmassa, jossa yritykset keräävät ja myyvät hyvinkin yksityiskohtaisia asiakastietoja kaupallisiin tarkoituksiin. Toisaalta on myös arvioitava, kuinka arvokasta tutkimuksellisesti olisi pelkkä tieto siitä, että käyttäjä on ladannut tietoja tietystä avoimen tiedon tietokannasta. Anonyymien käyttötietojenkin avulla voidaan selvittää käyttäjänäkökulmasta tärkeimpiä tai suosituimpia tietotyyppisiä. Suuret latausmäärät eivät välttämättä kerro vaikuttavuudesta.

Eräs olennainen kysymys, johon pelkkä tieto datan lataamisesta ei anna vastausta, on mihin hankittua tietoa käytetään vai käytetäänkö sitä mihinkään. Tämä on vaikuttavuustutkimuksen kannalta olennainen tieto. Tällaiseen tarkempaan datan käyttötietoon päästäisiin käsiksi parhaiten säännöllisesti tehtävien kyselytutkimuksien avulla. On kuitenkin huomattava, että avoin data muodostaa vain pienen osan eri käyttäjäryhmien hyödyntämästä datasta. Yritykset voivat käyttää sekä julkisen sektorin avointa dataa että muuta dataa (esim. yrityksen ostamat tiedot), ja tällöin avoimen datan ja muun datan vaikutusten erottaminen voi olla käytännössä mahdotonta. Datatalouden mekanismien ymmärtämiseksi tarvittaisiin tutkimustietoa yritysten datan käytöstä ja taloudellisista vaikutuksista laajemmin. Aihealueen aineistoanalyysiin perustuvan tutkimuksen mahdollistamiseksi tulisi kehittää datan käyttöä kuvastavaa tilastoaineistojen keräämistä. Datatutkimusta koskevia kysymyksiä tulisi lisätä Tilastokeskuksen säännöllisesti toteuttamiin lakisäätöihin kyselytutkimuksiin.

Keskeisten julkisen sektorin tietovarantojen suurin käyttäjäryhmä ovat yritykset: yritysten osuuden arvioitiin olevan keskimäärin liki puolet. Aineistoanalyysimme suomalaisyritysten joukossa osoittaa, että sekä avointa dataa että massadataa innovaatiotoiminnassaan hyödyntävät yritykset tekevät markkinoiden kannalta uusia tavara- tai palveluinnovaatioita suhteellisesti huomattavasti useammin kuin muut yritykset. Erityisen todennäköistä tämä on niiden dataa hyödyntävien yritysten keskuudessa, joissa digitaalisten tuotteiden merkitys liiketoiminnalle on huomattavaa. Koska innovaatiot ovat tuottavuuden ja kasvun keskeinen lähde, viittaavat tutkimustuloksemme mahdollisiin laajamittaisempiin datan hyödyntämisestä saataviin taloudellisiin hyötyihin.

Empiirinen tutkimuksemme osoittaa myös, että dataa uusien palveluiden tai tuotteiden kehittämisessä hyödyntäneet yritykset kasvoivat muita yrityksiä nopeammin joillakin tietointensiivisillä sektoreilla. Syksyllä 2011 tai sitä edeltävinä kolmena vuonna dataa innovaatioiden kehittämisessä käyttäneiden informaatio- ja viestintätoimialan yritysten liikevaihto kasvoi vuosina 2012–2014 keskimäärin yli 17 prosenttia enemmän kuin saman toimialan dataa innovaatioissa hyödyntämättömien yritysten liikevaihto. Julkista tietoa innovaatioiden kehittämisessä käyttäneiden yritysten liikevaihdon kasvu ei kuitenkaan poikennut selkeästi niiden yritysten kasvusta, jotka eivät käyttäneet julkista tietoa. *Yksittäisistä tietotyypeistä* vain liikennetietojen hyödyntäminen innovaatiotoiminnassa oli yhteydessä yritysten liikevaihdon kasvuun. Tutki-

mustuloksia tulkittaessa on huomattava, että useat merkittävät julkiset tietovarannot (esim. maastotiedot ja säätiedot) avattiin vasta vuoden 2011 jälkeen. Tämän takia tutkimuslöydöksistä ei voida vetää laajakantoisia johtopäätöksiä julkisen sektorin avoimen datan ja yritysten kasvun välisestä yhteydestä.

Lisätutkimusta, toimenpiteitä ja toimintaohjeita tarvitaan selventämään henkilötiedon, omadatan ja avoimen datan välistä suhdetta. Henkilötietoa koskee valtava määrä lakeja, ja EU:n tietosuoja-asetus vaikuttaa paitsi tietojen hallintaan organisaatioissa, myös avoimen datan hyödyntämiseen. Erityisesti henkilötietojen hyödyntäminen omadatan ja laajemmin anonymisoinnin kautta esimerkiksi tutkimuksessa ovat lisähuomiota vaativia kysymyksiä. Myös kansainvälistä työtä (esim. W3C – World Wide Web Consortium) datan laadun ja käytettävyyden parantamiseksi kannattaa seurata.

Kunnat ovat merkittävä kansallisten tietovarantojen tiedontuottaja. Kuntien rooli tietovarantojen hallinnassa ja datan avaamisessa on kuitenkin rajattu tämän hankkeen tutkimuskysymysten ulkopuolelle. Erityisesti Suomen kuudessa suurimmassa kaupungissa on avattu dataa ja tehty yhteistyötä avoimen datan hallinnan harmonisoimiseksi ja käytön edistämiseksi. Tästä suurten kaupunkien 6Aika-yhteishankkeesta kertynyttä tietoa ja osaamista olisi tärkeää levittää muihin kuntiin, ja ylipäätään tutkia kuntien osuutta niin avoimen datan tuottajina, julkaisijoina kuin käyttäjinäkin.

Julkisen datan avaamisen eturintamassa pysymisestä voi olla Suomelle konkreettisia hyötyjä. Suomi on sopivan pienen markkinaan kokeilla uusia asioita, joita voidaan myöhemmin skaalata suuremmille markkinoille. Parhaimmillaan edelläkävijyys mahdollistaa uusien vientiin suunnattavien toimintamallien ja tuotteiden kehittämisen. Se voi myös lisätä suomalaisten asiantuntijoiden painoarvoa kansainvälisessä yhteistyössä avoimen datan infrastruktuuria ja erilaisia standardeja ja suosituksia kehitettäessä sekä helpottaa ulkomaalaisten resurssien (esimerkiksi investoinnit ja osaaminen) saamista Suomeen. Kansainvälisyydessä olisi hyvä huomioida esimerkiksi Open Government Partnership -kumppanuusohjelma ja Open Data Charter⁴⁴ -julistus sekä kansainväliset avoimen datan vertailut kuten Open Data Barometer⁴⁵ ja Global Open Data Index⁴⁶, jotka kaikki auttavat kehittämään ja ylläpitämään laadukasta toimintaa.

⁴⁴ <http://opendatacharter.net/>

⁴⁵ <http://opendatabarometer.org/>

⁴⁶ <http://index.okfn.org/>

LÄHDELUETTELO

Antikainen, J. & Eskelinen, J. & Koski, H. & Niemi, T. & Pajarinen, M. & Pyykkönen, S. & Vries, M. (2016). Massadatasta liiketoimintaa ja tehokkaita julkisia palveluita. Valtioneuvoston selvitys ja tutkimustoiminnan julkaisusarja 16/2016. Saatavissa: <http://tietokayttoon.fi/julkaisu?pubid=10701>. Viitattu 12.1.2017.

Bensoussan, B. E. & Fleisher, C. S. (2008). Analysis without paralysis. 10 tools to make better strategic decisions. FT Press, Yhdysvallat. 228 ss.

Berends, J. & Carrara, W. & Engbers, W & Vollers, H. (2017). Re-Using Open Data – A Study On Companies Transforming Open Data Into Economic & Societal Value. Saatavissa: https://www.europeandataportal.eu/sites/default/files/re-using_open_data.pdf. Viitattu 26.2.2017.

Berners-Lee, T. (2009). Linked Data. WWW-dokumentti. Saatavissa: <https://www.w3.org/DesignIssues/LinkedData.html>. Viitattu 1.3.2017.

Boston Consulting Group (2012). The Value of Our Digital Identity. Liberty global Policy Series. PDF-dokumentti. Saatavissa: <http://www.libertyglobal.com/PDF/public-policy/The-Value-of-Our-Digital-Identity.pdf>. Viitattu 2.1.2017.

Economist (2013). Open Data: A New Goldmine. WWW-dokumentti. Saatavissa: <http://www.economist.com/news/business/21578084-making-official-data-public-could-spur-lots-innovation-new-goldmine>. Viitattu 16.3.2016

Economist (2015). Open Government Data: Out of the Box. WWW-dokumentti. Saatavissa: <http://www.economist.com/news/international/21678833-open-data-revolution-has-not-lived-up-expectations-it-only-getting>. Viitattu 16.3.2016.

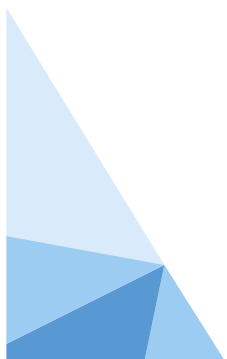
Euroopan komissio (2017). Euroopan unionin avoimen datan portaali. WWW-dokumentti. Saatavissa: <https://data.europa.eu/euodp/fi/about>. Viitattu 27.1.2017.

Green, B. & Cunningham, G. & Ekblaw, A. & Kominers, P. & Linzer, A. & Crawford, S. (2017). Open Data Privacy Playbook. Berkman Klein Center for Internet & Society. Harvardin yliopisto. PDF-dokumentti. Saatavissa: <https://cyber.harvard.edu/publications/2017/02/opendataprivacyplaybook>. Viitattu 27.2.2017.

Greengard, S. (2015). The Internet of Things. The Mit Press Essential Knowledge Series. MIT Press, Yhdysvallat. 210 ss.

Kauhanen-Simanainen, A. & Suurhasko, M. (2015). Avoimesta datasta innovatiiviseen tiedon hyödyntämiseen. Valtiovarainministeriön julkaisuja 31/2015. PDF-dokumentti. Saatavissa: <http://vm.fi/documents/10623/1107406/Avoimen+tiedon+ohjelman+loppuraportti/>. Viitattu 30.9.2016.

Koski, H. (2010). Julkisen tiedon hinnoittelu ja tarjonta Suomessa. Valtioneuvoston kanslia. Kasvutyöryhmän taustamuistio.



Koski, H. (2011). Does marginal cost pricing of public sector information spur firm growth? Etna Discussion Paper No. 1260.

Koski, H. (2015). Avoimen tiedon vaikuttavuus – esitutkimus. Valtiovarainministeriön julkaisu- ja 15a/2015.

McKinsey Global Institute (2011). Big Data. The Next Frontier for Innovation, Competition and Productivity. PDF-dokumentti. Saatavissa: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation. Viitattu 10.10.2016.

Moody, D. & Walsh, P. (1999). Measuring The Value Of Information: An Asset Valuation Approach. Seventh European Conference on Information Systems (ECIS'99). 17 ss.

Nurmi, P. (2017). EU:n yleisen tietosuoja-asetuksen Suomen malli. Oikeusministeriö. WWW-dokumentti. Saatavissa: <http://oikeusministerio.fi/fi/index/blogi/6Km6t6RUd/2017/vry1Vd3ix.html.stx>. Viitattu 20.2.2017.

Open Knowledge Finland (2015). Avoimen rajapinnan (API) määritelmä ja periaatteet suunniteluun. WWW-dokumentti. Saatavissa: <http://avoinrajapinta.fi/>. Viitattu 1.3.2017.

Open Knowledge International (2017). Lightweight standards and tooling to make it effortless to get, share, and validate data. WWW-dokumentti. Saatavissa: <http://frictionlessdata.io>. Viitattu 1.3.2017.

OpenDataSoft (2017). A List of 2600+ Open Data Portals Around the World. WWW-dokumentti. Saatavissa: <https://www.opendatasoft.com/a-comprehensive-list-of-all-open-data-portals-around-the-world/>. Viitattu 1.3.2017.

Paikkatietoikkuna (2016). Maanmittauslaitos. WWW-dokumentti. Saatavissa: <http://www.paikkatietoikkuna.fi/web/fi/inspire-direktiivi>. Viitattu 9.1.2016.

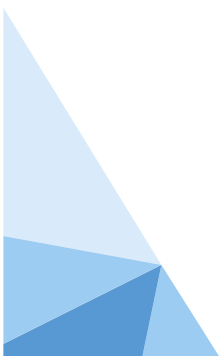
Parkkisenniemi, H. (2017). #Liikennelabra goes avoin data. Liikenneturvallisuusvirasto Trafi. PDF-dokumentti. Saatavissa: https://www.trafi.fi/filebank/a/1485324627/96fab34613da4bc-14c2cb809f1248df8/23911-Trafi_Liikennelabra_goes_avoin_data_24_1_2016_Hannu_Parkkisenniemi.pdf. Viitattu 30.1.2017.

Patentti- ja rekisterihallitus (2016). Vuosikertomus 2015. PDF-dokumentti. Saatavissa: https://www.prh.fi/stc/attachments/PRH_Vuosikertomus_2015_fin_WWW.pdf. Viitattu 1.1.2017.

Patentti- ja rekisterihallitus (2017). Tietoa avoimen datan rajapinnoista. WWW-dokumentti. Saatavissa: <http://avoindata.prh.fi/index.html>. Viitattu 19.1.2017.

Pham, N. (2011). The Economic Benefits of Commercial GPS Use in the U.S. and The Costs of Potential Disruption. PDF-dokumentti. Saatavissa: <http://www.gpsworld.com/wp-content/uploads/2013/03/GPS-Report-June-22-2011.pdf>. Viitattu 20.1.2017.

Poikola, A. & Kola, P. & Hintikka, H. (2010). Julkinen data – johdatus tietovarantojen avaamiseen. Liikenne- ja viestintäministeriö. PDF-dokumentti. Saatavissa: <https://www.scribd.com/doc/28845102/Julkinen-data>. Viitattu 20.2.2017.



Poikola, A. & Kuikkaniemi, K. & Kuittinen, O. (2014). My Data – johdatus ihmiskeskeiseen henkilötiedon hyödyntämiseen. Open Knowledge Finland. PDF-dokumentti. Saatavissa: <http://julkaisut.valtioneuvosto.fi/handle/10024/77875>. Viitattu 9.1.2017.

Reinvent Albany (2014). Listening to FOIL. Using FOIL Logs to Guide the Publication of Open Data. PDF-dokumentti. Saatavissa: <http://reinventalbany.org/wp-content/uploads/2014/07/Final-DEC-FOIL-Analysis.pdf>. Viitattu 12.2.2017.

Sasse, T. & Smith, A. & Broad, E. & Tennison, J. & Wells, P. & Atz, U. & Carrara, W. & Vollers, H. (2017). Recommendations for Open Data Portals: From Setup to Sustainability. European Unioni. PDF-dokumentti. Saatavissa: https://www.europeandataportal.eu/sites/default/files/edp_s3wp4_sustainability_recommendations.pdf. Viitattu 25.2.2017.

Stowe, M. (2015). Undisturbed REST: a Guide to Designing the Perfect API. lulu.com. 200 ss.

Talus, A. & Autio, E. (2017). Miten valmistautua EU:n tietosuoja-asetukseen? Selvityksiä ja ohjeita. Oikeusministeriö. PDF-dokumentti. Saatavissa: http://www.oikeusministerio.fi/fi/index/julkaisut/julkaisuarkisto/1487252333714/Files/OMSO_04_2017_OM_TSV_EU_tietosuoja_2.pdf. Viitattu 19.2.2017.

Tietosuojavaltuutetun toimisto (2012). Laadi tietotilin päätös. PDF-dokumentti. Saatavissa: http://www.tietosuoja.fi/material/attachments/tietosuojavaltuutettu/tietosuojavaltuutetuntoimisto/oppaat/6JfpzNVCh/Laadi_tietotilinpaatos.pdf. Viitattu 10.1.2017.

Trafi (2016). Tietotilin päätös 2015. PDF-dokumentti. Saatavissa: https://www.trafi.fi/filebank/a/1463394812/9478dbd63f9d555bc3cd6f5b5bc99e54/20667-Trafin_julkaisu_14-2016_-_Tietotilinpaatos_2015.pdf. Viitattu 10.1.2017.

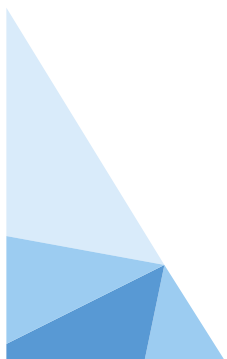
Valtioneuvosto (2011). Valtioneuvoston periaatepäätös julkisen sektorin digitaalisten tietoa-ineistojen saatavuuden parantamisesta ja uudelleen käytön edistämisestä. PDF-dokumentti. Saatavissa: <http://www.lvm.fi/documents/20181/782861/Ehdotus+valtioneuvoston+periaatepaatokseksi+-+Julkinen+tietoaineisto+%283.3.2011%29.pdf/50cd10fc-824b-407c-aa-be-2f56e333e108?version=1.0>. Viitattu 10.1.2017.

Valtioneuvosto (2016). Valtioneuvoston periaatepäätös datan hyödyntämisestä liiketoiminnassa. PDF-dokumentti. Saatavissa: <http://valtioneuvosto.fi/delegate/file/14723>. Viitattu 10.1.2017.

Verhulst, S. & Young, A. (2016). Open Data Impact When Demand and Supply Meet. PDF-tiedosto. Saatavissa: <http://odimpact.org/static/files/open-data-impact-key-findings.pdf>. Viitattu 16.3.2016.

Vickery, G. (2011). Review of Recent Studies on PSI Re-Use and Related Market Developments. PDF-dokumentti. Saatavissa: https://www.nsgic.org/public_resources/Vickery.pdf. Viitattu 20.10.2016.

Wiiala, N. & Kankaanrinne, J. (2016). PERTTI Perustietovarantojen tietomallien yhteensovittaminen. Julkisen hallinnon tietohallinnon neuvottelukunta (JUHTA). PDF-tiedosto. Saatavissa: <https://wiki.julkict.fi/julkict/juhta/juhta-n-jaostot/perustietovaranto-jaosto/perus->



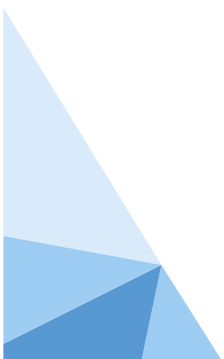
rekisterien-tietomallien-yhteensovittaminen/tyon-tulokset/pertti-selvityksen-loppuraportti-v1-29-1.2016/at_download/file. Viitattu 16.3.2016.

World Economic Forum (2013). Unlocking the Value of Personal Data: From Collection to Usage. PDF-dokumentti. Saatavissa: http://www3.weforum.org/docs/WEF_IT_UnlockingValuePersonalData_CollectionUsage_Report_2013.pdf. Viitattu 1.1.2017.

Wyns, B. & Bargiotti, L. & Loozen, N. & Loutas, N. & Dekkers, M. & De Keyser, M. & Goedertier, S. (2013). Good Practices for Identifying High Value Datasets and Engaging with Re-users: the Case of Public Tendering data. PDF-dokumentti. Saatavissa: https://www.w3.org/2013/share-psi/wiki/images/3/31/Share-PSI_Submission_Paper-PwC_v0.03.pdf. Viitattu 10.11.2016.

Väestörekisterikeskus (2011). Perustietovarannot – tietohuoltomme kansallispääoma. Julkisen hallinnon tietohallinnon neuvottelukunta (JUHTA). PDF-dokumentti. Saatavissa: https://wiki.julkict.fi/julkict/juhta/juhta-n-jaostot/perustietovaranto-jaosto/pertiva-kauden-2010-2012-dokumentteja/perustietovarannot-julkaisu-2011/perustietovarannot-2013-tietohuoltomme-kansallispaaoma/at_download/file. Viitattu 16.3.2016.

Young, A. & Verhulst, S. (2016). The Global Impact of Open Data. PDF-tiedosto. Saatavissa: <http://www.oreilly.com/data/free/the-global-impact-of-open-data.csp> Viitattu 5.2.2017.



LIITE 1 KESKEISET KÄSITTEET

6Aika on Suomen kuuden suurimman kaupungin (Helsinki, Espoo, Vantaa, Oulu, Tampere ja Turku) kestävä kaupunkikehityksen strategia, jonka aiheena ovat avoimet ja älykkäät palvelut. Strategiassa toteutetaan projekteja aina useamman kaupungin alueella ja sen kokonaisbudjetin arvioidaan kasvavan vuosina 2014-2020 lähes 100 miljoonaan euroon. Strategia painottuu avoimien innovaatioalustojen, avoimen datan ja rajapintojen ja avoimen osallisuuden ja asiakkuuden hankkeisiin. Lisätietoja: <https://6aika.fi/>.

Avoin data (open data) on dataa, joka on kenen tahansa uudelleen käytettävissä maksutta, luvallisesti ja koneluettavassa muodossa. Avointa dataa voidaan vapaasti laajentaa, kiertää ja yhdistellä edelleen eri sovelluksissa. Esimerkiksi Tilastokeskuksen väestötalastot, Ilmatieteen laitoksen sää- ja tutkahavainnot ja Maanmittauslaitoksen maastotietokannat ovat avointa dataa.

Avoin lisenssi (open license) on käyttöluupa, joka sallii aineiston vapaan levittämisen, muokkauksen ja käytön, myös kaupalliseen tarkoitukseen.

Avoin ohjelmointirajapinta (avoin API) on ohjelmointirajapinta, jonka kaikki ominaisuudet ovat julkisia ja jota voi käyttää ilman rajoittavia ehtoja. Open Knowledge Finland ry, Coss ry, valtiovarainministeriö, Helsinki, Espoo, Vantaa, Turku, Tampere ja Oulu ovat olleet mukana laatimassa yhteistä avoimen ohjelmointirajapinnan määritelmää, joka on julkaistu osoitteessa: <http://avoinrajapinta.fi/>.

Avoin tieto on tietoa, jota kuka tahansa voi eri tavoin käyttää uudelleen maksutta ja luvallisesti.

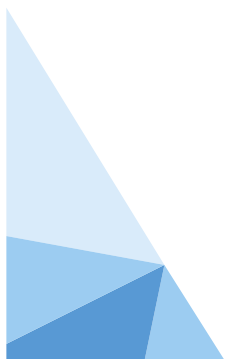
CC BY 4.0 eli Creative Commons Nimeä 4.0 on avoimelle datalle suositeltu lisenssi (JHS 189). Lisenssin tarkoituksena on antaa oikeudellinen varmuus siitä, että käyttöluvan kohteena olevaa dataa voi vapaasti käyttää kaikin mahdollisin tavoin edellyttäen, että datan lähde mainitaan käyttöluvasta ilmenevällä tavalla.

CC0 on merkintä, joka käytännössä ilmaisee, että julkaisija luopuu kaikista yksinoikeuksistaan, joista hän lain mukaan voi luopua (sisältyy JHS 189). CC0 on suositeltava lisenssi kaikelle metadatalle. Voidaan käyttää, jos aineisto on tekijänoikeuden tai lähioikeuksien suojan piirissä (edes osittain).

CKAN on avoimella lähdekoodilla rakennettu sisällönhallintaohjelmisto, joka toimii avoimen datan jakeluportaalina. Se tarjoaa työkaluja datan julkaisemiseen, jakamiseen, löydettävyyteen ja käyttöön. Ohjelmistoa käyttävät organisaatiot, kunnat, valtiot ja yritykset, jotka haluavat avata ja julkaista avointa dataa.

Creative Commons -lisenssit ovat maailmanlaajuinen standardi avoimen tiedon julkaisuun. Monissa maissa tekijänoikeuslaki voi suojella pelkkiä datataulukoitakin, joten avattavaan dataan on aina liitettävä avoin lisenssi.

Dataformaatti (data format) on tiedostomuoto, johon dataa tallennetaan. Esimerkiksi CSV, JSON ja XML.



Datapolitiikka on organisaation tietoaaineistojen tehokäyttöä edustava johtamisen ja käytännön työskentelyn väline. Datapolitiikka määrittelee painopisteet, joilla organisaation tietovarantojen käyttöä edistetään.

EU GDPR on uusi kaikkia EU-maita koskeva yleinen tietosuoja-asetus, joka astuu voimaan 25.5.2018. Tietosuoja-asetuksessa on laaja kansallinen liikkumavara ja sen tarkka vaikutus Suomen lainsäädäntöön tarkentuu myöhemmin. Suomessa henkilötietoja käsitellään noin 600 eri laissa ja asetuksessa.

Hackathon tulee sanojen ”hack” ja ”marathon” yhdistelmästä, jossa ”hack” tarkoittaa eksploraatiivista ohjelmointia (uuden etsimistä & kokeilemistä) ja marathon kuvaa intensiivistä ja tehokasta toimintaa. Tyypillisesti päivästä viikkoon kestävä tapahtuma, jonka aikana joukko ihmisiä suunnittelee ja toteuttaa tiiminä erilaisia ideoita, ohjelmistoja tai sovelluksia.

Henkilötieto on kaikenlaisia luonnollista henkilöä taikka hänen ominaisuuksiaan tai elinolosuhteitaan kuvaavia merkintöjä, jotka voidaan tunnistaa häntä tai hänen perhettään tai hänen kanssaan yhteisessä taloudessa eläviä koskeviksi. (Henkilötietolaki 22.4.1999 / 523, 3 § määritelmät).

Innovatiiviset palvelut ovat palveluita, jotka hyödyntävät tietoa ja osaamista kaupallisesti tai yhteiskunnallisesti uudella tavalla.

INSPIRE on EU-direktiivi, jonka avulla kansallisista paikkatietoaaineistoista ja -palveluista luodaan EU:n jäsenmaiden yhteinen, yhtenäinen ja helposti hyödynnettävä paikkatietoinfrastruktuuri. (Paikkatietoikkuna, 2016).

Julkisuuslaki on vuonna 1999 voimaan tullut laki ja asetus viranomaistoiminnan julkisuudesta (1999/621). Julkisuuslaki eli laki viranomaisten toiminnan julkisuudesta (621/1999, lyhenne JulkL, on yleissäännös, joka käsittelee viranomaisten hallussa olevien asiakirjojen julkisuutta, salaamista ja julkistamista. Perussäännön mukaan asiakirja on julkinen, ellei sen salaamiseen ole laissa erikseen mainittua syytä.

Kapa-laki on laki hallinnon yhteisistä sähköisen asioinnin tukipalveluista (571/2016). Laissa säädetään mm. julkisen hallinnon yhteisistä sähköisen asioinnin tukipalveluista, niitä koskevista vaatimuksista, niiden tuottamiseen liittyvistä tehtävistä sekä tuottamiseen liittyvästä henkilö- ja muiden tietojen käsittelystä. Taustalla ovat Kansallisen palveluarkkitehtuurin toteuttamisohjelmassa (KaPA) kehitettävät digitaalisten palvelujen infrastruktuuria edistävät tukipalvelut ja nykyisten sähköisen asioinnin tukipalveluiden tuotantovastuiden tarkennus palveluarkkitehtuurin mukaisiksi. Lisätietoja: <http://vm.fi/kapa-laki>.

Kapa-ohjelman (Kansallinen Palveluarkkitehtuuri) tavoitteena on tuottaa yleiskäyttöisiä ohjelmistokomponentteja digitalisaation tueksi. Näitä komponentteja ovat mm. kansallinen palveluväylä, palvelunäkymät, kansallinen tunnistusratkaisu sekä rooli- ja valtuutuspalvelu.

Koneluettavuus on datan ominaisuus, joka tukee datan koneellista yksilöimistä, tunnistamista ja poimimista.

Kumppanikoodari on organisaatioon palkattava ”ulkoinen” ohjelmointitaitoinen henkilö, jonka tehtävänä on kehittää työyhteisön sähköisiä palveluita. Ulkopuolinen objektiivinen näkökulma ja pyrkimys asioiden kyseenalaistamiseen mahdollistavat asioiden tarkastelun uudella tavalla. Kumppanikoodari voi kehittää avoimen datan avaamista ja hyödyntämistä organisaatio-

tiossa. Helsingin kaupungilla on kumppanikoodareita. Kansainvälisesti ideaa on sovellettu muun muassa Yhdysvalloissa 18F-yksikössä, jonka tehtävänä on digitalisoida käyttäjälähtöisesti ja kokeilemalla nopeasti julkisen hallinnon palveluita.

Käyttöehdot ovat ehdot, jotka koskevat tekijänoikeuden ja lähioikeuksien suojaaman aineiston käyttöä.

Käyttölupa (tai lisenssi) on käyttöehtoihin perustuva ei-yksinomainen oikeus käyttää aineistoa.

Latauspalvelu on verkkosivu tai -palvelu, jonka kautta voi ladata ("kopioida") tietyt avoimet tiedostot paikallisesti toiselle koneelle tai palvelimelle (ks. myös rajapintapalvelu).

Lisenssi kuvaa käyttöluvan eli käyttöehtoihin perustuvan oikeuden käyttää aineistoa.

Lynet on vuodesta 2008 toiminut luonnonvara- ja ympäristöalan tutkimusverkosto, joka kokoaa tutkijat yhteen tuottamaan tietoa ja ratkaisuja alan ongelmiin. Verkostoon kuuluvat Evira, Geologian tutkimuslaitos (GTK), Ilmatieteen laitos, Luke, MML, Syke ja VTT. Lisätietoja: <http://www.lynet.fi/>.

Maksuperustelaki (Valtion maksuperustelaki 21.2.1992/150) Maksuperustelaki on valtion viranomaisten suoritteista perittäviä maksuja koskeva yleislaki, jonka on ollut tarkoitus soveltaa mahdollisimman monien viranomaisten toimintaan. Lain ja sen nojalla annettavien säädösten ja määräysten perusteella viranomaiset voivat periä maksuja virkatoimista, palveluista, tuottamista tavaroista ja muusta toiminnastaan.

Massadata (big data) tai suuraineisto on dataa, jota on paljon, jota tulee nopeasti lisää ja joka on muodoltaan vaihtelevaa. Massadatan käsittely on vaikeaa perinteisillä tietojenkäsittelysovelluksilla. Laitteilta ja sovelluksilta vaaditaan valtavaa tallennuskapasiteettia ja suorituskykyä, jotta ne pystyisivät käsittelemään massadataa. Massadataa pyritään keräämään, säilyttämään ja analysoimaan merkityksellisen informaation löytämiseksi. Massadataa voidaan hyödyntää esimerkiksi tieteellisessä tutkimuksessa, tuotekehityksessä, markkinoinnissa ja päätöksenteossa.

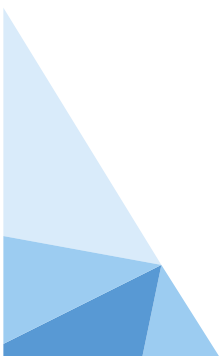
Metatieto (metadata) tarkoittaa tietoa tiedosta. Kuvailevaa ja määrittävää tietoa jostakin tietovarannosta tai sisältöyksiköstä. Esimerkiksi tekstidokumentin tiedot (viimeksi tallennettu, omistaja, versio, sijainti, julkaisupäivämäärä jne.) ovat metatietoa.

Mydata on toinen nimitys omadatalle. Omadata on "MyData"-termin suomennus.

Nettobudjetoitu virasto tarkoittaa virastoa, joka rahoittaa toimintansa pääosin tai kokonaan omilla palvelutuloillaan.

Ohjelmointirajapinta (Application Programming Interface, API) määrittelee, miten ohjelmisto tarjoaa tietoja tai palveluita sovelluksille tai muille tietojärjestelmille.

Omadata on ihmiskeskeinen lähestymistapa henkilötiedon hallintaan ja käsittelyyn. Siinä ihmisille annetaan oikeus ja pääsy heistä kerättyyn dataan kuten ostotietoihin, liikennetietoihin, teletietoihin, terveystietoihin, taloustietoihin ja eri verkkopalveluihin kertyvään dataan. (Poikola et al., 2014, s. 9).



Paikkatieto on tietoa kohteista, joiden sijainti Maan suhteen tunnetaan.

Perustietovarantoja ovat sekä laissa säädetyt perusrekisterit että muut keskeiset kansallisen tason tietovarannot. Perusrekisterillä tarkoitetaan keskitettyä tietojärjestelmää, johon on kattavasti ja mahdollisimman virheettöminä kerätty keskeisiä tietoja yhteiskunnan perusyksiköistä, kuten esimerkiksi henkilöistä tai kiinteistöistä.

Perusrekisteri on keskitetty tietojärjestelmä, johon on kattavasti ja mahdollisimman virheettöminä kerätty keskeisiä tietoja yhteiskunnan perusyksiköistä. Suomessa perusrekistereitä ovat väestötietojärjestelmä (VTJ), kiinteistötietojärjestelmä (KTJ) sekä yritys- ja yhteisötietojärjestelmä (YTJ).

PSI-direktiivi on direktiivi julkisen sektorin hallussa olevien tietojen uudelleenkäytöstä (Euroopan parlamentin ja neuvoston direktiivi 2003/98/EY).

QGIS on avoimen lähdekoodin paikkatietojen käsittelyyn käytettävä ohjelmisto, jota käytetään datan katseluun, muokkaukseen ja analysointiin.

R on avoimen lähdekoodin ohjelmointikieli, joka on kehitetty erityisesti tilastolliseen laskentaan ja grafiikan tuottamiseen. Sitä hyödynnetään tilastojen käsittelyssä ja data-analyysissä laajasti.

Raakadata on koneellisesti luettavissa oleva data, johon ei kohdistu tekijänoikeuksia.

Rajapintapalvelu tekninen käyttöyhteys, jonka avulla asiakassovellus ottaa yhteyttä palvelimelle ja noutaa sieltä dataa määritteiden mukaan. Rajapintapalveluiden käyttö edellyttää, että käytössä on ohjelmisto, joka tekee pyynnöt ko. palvelimelle.

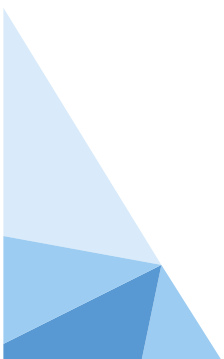
Rakenteinen aineisto (structured data) on sähköinen aineisto, jossa aineiston osat ja niiden väliset suhteet ilmaistaan sovitulla tavalla.

Soterko on sosiaali- ja terveysalan asiantuntijalaitosten yhteenliittymä, joka aloitti toimintansa vuonna 2011. Siinä yhteistyökumppaneina ovat mm. THL, TTL, STUK, Evira, Fimea, Ilmatieteen laitos, Kela, Syke, Vatt ja VTT. Yhteenliittymällä on seitsemän tutkimus- ja kehittämisohjelmaa sekä tietovarantojen kehittämisohjelma. Lisätietoja: www.soterko.fi.

Tiedonhallinnalla tarkoitetaan toimenpiteitä, menetelmiä ja käytäntöjä, jotka mahdollistavat tietojen löytämisen, käsittelyn ja hyödyntämisen tietojen koko elinkaaren ajan.

Tietoaineisto (dataset; data set) on tietovälineiden ja niihin tallennettujen tietojen muodostama kokonaisuus. Esimerkiksi yksittäinen taulukko tai luettelo.

Tietojen uudelleenkäyttö tarkoittaa sitä, että luonnolliset henkilöt tai oikeushenkilöt käyttävät julkisen sektorin elinten hallussa olevia asiakirjoja kaupallisiin tai muihin kuin kaupallisiin tarkoituksiin, jotka poikkeavat alkuperäisestä julkisesta tehtävästä, jota varten asiakirjat tuotettiin (EU-direktiivi 2003/98/EY). Tietojen uudelleenkäyttö voi käytännössä olla esimerkiksi tietojen yhdistämistä toisiin tieto- tai muihin sisältöihin tai kokonaan uusien palvelujen luomista tietojen pohjalta.



Tietotilinpäätös on raportti, joka syntyy organisaation sisäisen tarkastelun tuloksena ja antaa kokonaiskuvan organisaation tietojenkäsittelyn nykytilasta. Tietotilinpäätös kuvaa myös henkilötietolain mukaisen hyvän tietojenkäsittelytavan noudattamista.

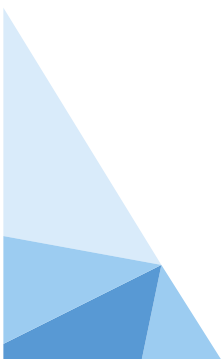
Tietotuote on yhden tai useamman tietoaineiston tai tietoaineistosarjan tiedoista koottu tuote, joka voidaan luovuttaa sellaisenaan tai sisällön perusteella rajattuna tiettyyn käyttötarkoitukseen.

Tietovaranto (information resource; information pool) on looginen tietoaineistojen kokonaisuus, jota jokin taho hallinnoi. Toiminnan tarpeista johdettu ja hallinnollisista syistä määritelty tietojen kokonaisuus. Se voi koostua yhden tai useamman järjestelmän sisältämistä tiedoista.

Tietovarantojen avaaminen tarkoittaa julkisten tietojen tarjoamista koneluettavassa muodossa vapaasti uudelleen käytettäväksi. Tietovarantojen avaaminen uudelleenkäyttöä varten ei koske salassa pidettäviä tietoja eikä tietoja, joiden käyttö on lainsäädännön perusteella rajoitettua. Tietovarantoja voidaan rajatusti avata myös joidenkin ryhmien, kuten tutkijoiden tai toisten viranomaisten käyttöön. Tällöin ei ole kyse avoimesta datasta sen varsinaisessa merkityksessä.

Tulosopimus on viraston ja ministeriöiden välinen sopimus alkavan vuoden toiminnallisista tavoitteista myönnettyjen määrärahojen puitteissa, tyypillisesti tehty neljän vuoden ajanjaksolle.

Viranomaistieto Julkisen hallinnon viranomaisen tehtäviään varten keräämä tai kokoama tieto.



LIITE 2 ETLAN KYSELYLOMAKE REKISTERINPITÄJILLE

Kysely tietotuotteiden ja avoimen datan hyödyntämisestä

JOHDANTO

Tietotuotteilla käsitetään tässä kyselyssä tuottamianne ja tarjoamianne yksittäisiä tietoja, aineistoja ja tietokantoja sähköisessä tai painetussa muodossa (esim. rekisteriotteet, tilastoaineistot, online-tietokannat). Tähän ei lueta tuottamianne julkaisuja ja raportteja.

Avoin data tarkoittaa keräämäännne tai tuottamaanne dataa, joka on konekielisenä sekä vapaasti ja veloituksettomasti kansalaisten ja kehittäjien käytettävissä, uudelleenkäytettävissä ja muokattavissa.

Vastauksenne ovat täysin luottamuksellisia eikä vastaajakohtaisia tietoja paljasteta ulkopuolisille. Kyselyyn vastaaminen kestää noin 15 minuuttia.

MAKSUTTOMIEN TIETOTUOTTEIDEN TARJONTA

K1. Tarjoatteko maksuttomia tietotuotteita? KYLLÄ/EI: JOS EI -> K4.

K2. Kuinka suuren prosenttiosuuden (0-100) maksuttomat tietotuotteet kattavat kaikista tarjoamistanne tietotuotteista, joiden avaamiselle ei ole lakisääteistä estettä?

K3. Kuinka suuri prosenttiosuus (0-100) tietotuotteistanne on saatavilla avoimena datana (tietotuotteet, joiden avaamiselle ei ole lakisääteistä estettä)?

MAKSULLISTEN TIETOTUOTTEIDEN TARJONTA

Irtiottokuluilla tarkoitetaan tiedon luovuttamisesta ja jakamisesta syntyviä välittömiä kustannuksia.

Omakustannehinnalla tarkoitetaan sitä, että maksun suuruus vastaa tietotuotteen tuottamisesta aiheutuvien kokonaiskustannusten määrää.

K4. Arvioika, kuinka suuri osa tarjoamistanne maksullista tietotuotteista on hinnoiteltu seuraavien luokitusten mukaisesti ja kuinka suuret tulot vuonna 2015 eri tavoin hinnoitelluista tietotuoterhyhmistä on saatu.

	Hinnoittelua käytetään	Tietotuotteiden %-osuus (0 -100) maksullisista tietotuotteista	Tulot (1000 euroa) vuonna 2015
Nimellinen/irtiottokuluihin perustuva hinta	Kyllä/Ei		
Omakustannehinta	Kyllä/Ei		
Liiketaloudellinen hinnoittelu	Kyllä/Ei		

K5. Tarjoatteko maksullisia tietotuotteitanne jollekin seuraavista tahoista erikseen määritellyin hinnoitteluperustein? Jos tarjoatte, niin mikä on hinnoitteluperuste? (voitte valita kaikki käytössä olevat)

	Tietotuotteita tarjotaan	Hinnoitteluperuste			
		Nimellinen hinta	Omakustannus-hinta	Liiketaloudellinen hinnoittelu	Ilmainen
Tutkimuskäyttö	Kyllä/ei				
Opetuskäyttö	Kyllä/ei				
Viranomais-käyttö	Kyllä/ei				
Yksityiskäyttö	Kyllä/ei				



K6. Kuinka suuret olivat tietotuotteistanne kertyneet tulot yhteensä vuonna 2015? (tuhatta euroa)

K7. Kuinka suuret olivat tietotuotteidenne kustannukset yhteensä vuonna 2015? (tuhatta euroa)

TÄRKEIMMÄT TIETOTUOTTEET

K8. Seuraavaksi kysymme kolmesta eniten käytetystä tietotuotteestanne tai tietotuoteryhmästäne. Kuvaillkaa kyseiset tietotuotteet, sekä niiden hinnoitteluperuste ja saatavuus alla oleviin kohtiin.

K8_1A. Tietotuotteen 1 kuvaus: AVOIN KENTTÄ

K8_1B. Tietotuotteen 1 hinnoittelu (useampi valinta sallittu):

1. Maksuton
2. Nimellinen/irtiottokustannuksiin perustuva
3. Omakustannehinta
4. Liiketaloudellinen hinnoittelu

K8_1C. Onko tietotuote 1 saatavissa avoimena datana, ts. konekielisenä sekä vapaasti ja veloituksettomasti kansalaisten ja kehittäjien käytettävissä, uudelleenkäytettävissä ja muokattavissa? KYLLÄ/EI

K8_2B. Tietotuotteen 2 kuvaus: AVOIN KENTTÄ

K8_2B. Tietotuotteen 2 hinnoittelu (useampi valinta sallittu):

1. Maksuton
2. Nimellinen/irtiottokustannuksiin perustuva
3. Omakustannehinta
4. Liiketaloudellinen hinnoittelu

K8_2C. Onko tietotuote 2 saatavissa avoimena datana, ts. konekielisenä sekä vapaasti ja veloituksettomasti kansalaisten ja kehittäjien käytettävissä, uudelleenkäytettävissä ja muokattavissa? KYLLÄ/EI

K8_3A. Tietotuotteen 3 kuvaus: AVOIN KENTTÄ

K8_3B. Tietotuotteen 3 hinnoittelu (useampi valinta sallittu):

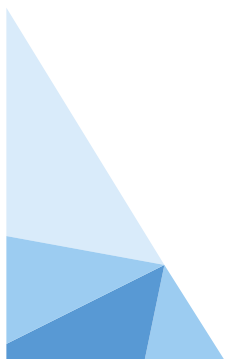
1. Maksuton
2. Nimellinen/irtiottokustannuksiin perustuva
3. Omakustannehinta
4. Liiketaloudellinen hinnoittelu

K8_3C. Onko tietotuote 3 saatavissa avoimena datana, ts. konekielisenä sekä vapaasti ja veloituksettomasti kansalaisten ja kehittäjien käytettävissä, uudelleenkäytettävissä ja muokattavissa? KYLLÄ/EI

TIETOTUOTTEIDEN KÄYTTÖ JA KEHITTÄMISMAHDOLLISUUDET

K9. Arvioikaa eri käyttäjäryhmien osuus (perustuen käytettyjen tietotuotteiden lukumäärään) omien tietotuotteidenne käytössä (0–100 %)

1. Yksityishenkilöt
2. Yritykset
3. Tutkimuslaitokset/yliopistot
4. Julkisen sektorin muut toimijat



K10. Mikäli käytätte omakustanne-/liiketaloudellista hinnoittelua, millä edellytyksillä näkisitte mahdollisena siirtyä käyttämään pelkästään nimellisiä/irtiottokustannuksiin perustuvia hintoja? Voitaisiko tietotuotteenne tarjota maksuttomina niin, että kustannukset katettaisiin budjettirahoitteisesti? Jos ei, niin mikä on esteenä? AVOIN KENTTÄ

K11. Minkälaisia keräämiänne tieto(kanto)ja tai aineistoja organisaationne ei tarjoa avoimena datana muista kuin lakisääteisistä syistä (esim. tietosuoja)? Mistä syistä/syistä näitä tietoja ei tarjota avoimena datana ja millä edellytyksillä olisi mahdollista avata tätä dataa? AVOIN KENTTÄ

K12. Onko teillä suunnitelmia julkaista ei-avoimia tietotuotteitanne avoimena datana? AVOIN KENTTÄ

K13. Kuinka monta prosenttia tietotuotteistanne, jotka eivät ole avoimia, mutta olisivat lainsäädännön puitteissa avattavissa, on (0–100 %)

1. Konekielisesti luettavissa
2. Verkossa saatavilla
3. Maksullisia

K14. Pystyttekö tunnistamaan käyttäjät, jotka hyödyntävät organisaationne avointa dataa? KYLLÄ/EI

K15. Pystyttekö erottelamaan missä määrin aineistojanne hyödyntävät yritykset tai yksityiset henkilöt? KYLLÄ/EI

K16. Pystyttekö identifioimaan käyttäjien henkilöllisyyden tai organisaation? KYLLÄ/EI

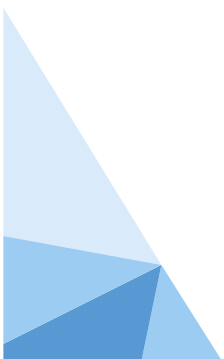
K17. Mikäli ette voi tunnistaa avoimen datan käyttäjiä, mikä on esteenä? AVOIN KENTTÄ

K18. Olisiko mahdollista identifioida käyttäjät (esim. tutkimuskäyttöä varten)? AVOIN KENTTÄ

K19. Mikäli haluatte antaa selvennyksen johonkin vastaukseen (tai useampaan vastaukseen), niin alla olevaan kenttään voi kirjoittaa lisähuomioita. AVOIN KENTTÄ

LOPETUS

Kiitokset vastauksistanne! Painakaa vielä 'lähetä vastaukset' painiketta tallentaaksenne vastauksenne.



LIITE 3 DATAN AVAAJIEN TEEMAHAASTATTELURUNKO

Osa-alue 0: Haastateltavien taustatiedot sekä tutkimuksen esittely

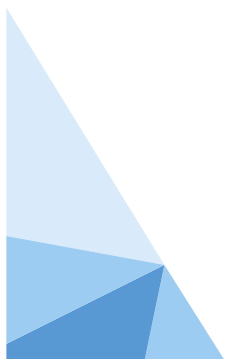
1. Haastateltavien nimi / nimet
2. Haastateltavien esittäytyminen
3. Haastateltavien työtehtävät organisaatiossa
4. Tutkimuksen lyhyt esittely sekä työnjako Etlan kanssa
5. Luvan kysyminen haastattelun äänittämiseen sanelimella

Osa-alue 1: Avoimen datan suunnitelmat

1. Onko organisaationne kartoittanut ja dokumentoinut hallussa olevat tietovarannot (tietokannat / tietovarastot)?
2. Miten arvioisitte näiden tietojen oikeellisuutta ja ajantasaisuutta?
3. Arvioikaa, miten paljon tietovarantojen sisällöt muuttuvat kuukaudessa. Jos ette tiedä tarkasti, esittäkää paras arvio asiasta.
4. Sisältääkö dokumentaatio metadatakuvaukset informaatioisällöistä ja niiden semanttisesta yhteensopivuudesta?
5. Onko organisaatiollanne voimassaoleva avoimen datan suunnitelma (dokumentoitu) tai vastaava? Milloin suunnitelma on päivitetty viimeksi? Mistä se löytyy?
6. Mitä tietopalveluihin ja tietovarantoihin liittyviä kehityssuunnitelmia tai -hankkeita on meneillään tai suunnitteilla?
7. Millainen valmius/kyvykyys/suunnitelma organisaatiollanne on uutta EU-tietosuoja-asetusta varten?
8. Miten PSI-direktiivi (Public Sector Information) vaikuttaa toimintaanne? Millaisella aikataululla?
9. Mitkä muut ajankohtaiset lainsäädäntö- tms. hankkeet vaikuttavat toimintaanne nyt?
10. Oletteko perehtyneet Avoindata.fi -ohjeistukseen (<https://www.avoindata.fi/fi/opas>)? Miten arvioitte ko. ohjeistuksen ja sen hyödyn, sekä mahdolliset puutteet/kehitystarpeet? Miten kehittäisitte ohjeistusta?

Osa-alue 2: Tietovarantojen ja avoimen datan hallinta ja organisointi

1. Julkaisetteko avointa dataa ohjelmistorajapinnan kautta ja/tai tiedostona? Millaisia tulevaisuuden suunnitelmia avoimen datan julkaisemiseen organisaatiossanne liittyy?
2. Missä palveluissa tiedostoja/tietoa julkaistaan (omat kotisivut, avoindata.fi, API:t, EU:n portaalit / joku muu, mikä _____?)
3. Miten usein päivitätte avoimena datana julkaistua tietoa? Onko erilaisia päivitystahteja eri datoilte?
4. Onko avoimen datan julkaisunne jatkuvaa vai projektimaista? Kuvailkaa toimintaanne
5. Miten kauan datasetin julkaisemisprosessi tyypillisesti kestää ja millaista resurssointia se tarvitsee?
6. Tukevatko tietojärjestelmät avoimen datan julkaisua ja kuinka paljon asia vaatii ns. ”käsityötä”?
7. Arvioikaa, kuinka monta virkamiestä organisaatiossanne on perehtynyt avoimeen dataan?
8. Voitaisiinko joitain osia perusrekistereistänne / tietovarannoistanne avata? Mitä osia?



Osa-alue 3: Tietoaineistojen käyttö ja hyödyntäjät

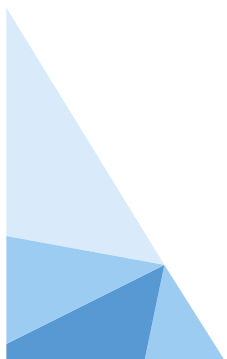
1. Miten tietoaineistojen/-varantojen käyttöä seurataan? Millaisia aineistoja tästä kertyy, miten niitä analysoidaan ja kuinka kauan niitä säilytetään?
2. Keitä ovat teidän perusrekisterien/avoimen datan/tietovarantojen tärkeimmät käyttäjäryhmät? Esittäkää paras ja valistunein arvio, jos ette tiedä tarkasti.
3. Onko käyttäjät tiedossa vielä tarkemmin yksilöityinä?
4. Onko tiedossa potentiaalisia käyttäjiä, joita ei jostain syystä pystytä palvelemaan? Miksi?
5. Miten arvioitte tietoaineistojen arvoa? Käyttekö jotain liiketoiminnallista analysointimenetelmiä avoimen datan avaamisen tukena (esimerkiksi kustannushyöty-analyysi)?
6. Miten priorisoitte avoimen datan työtä / datan avaamista?

Osa-alue 4: Budjetti (taloudellinen näkökulma)

1. Miten hahmotatte tai määrittelisitte tietovarantojenne tuottaman arvon ja taloudellisen merkityksen? Miten mittaatte sitä? Miten asiaa johdetaan organisaatiossanne?
Yleisesti informaation käsittelyyn käytetään organisaatioissa yhä enemmän resursseja, mutta niiden arvo ei yleensä taloudellisessa raportoinnissa (esimerkiksi tuloslaskelma ja tase).
2. Mikä on maksullisten tietotuotteiden budjettimerkitys organisaatiollenne? (*"kerro lisää verrattuna ETLA-tutkimukseen"*)
3. Miten arvioitte väitettä: *"dataa myymällä julkinen sektori käy pääasiassa kauppaa keskenään"*? Osaatteko arvioida kuinka paljon yritykset / kuluttajat ostavat tietoaineistoja? Mitkä yritykset ostavat ja mitä?
4. Mitkä ovat tieto- ja tietotuotteitten hinnoitteluperiaatteet ja -käytännöt? Milloin ja miten tietotuotteenne on hinnoiteltu? Millä logiikalla, miten tätä hinnoittelua seurataan/tarkistetaan? (*"kerro lisää verrattuna ETLA-tutkimukseen"*)
5. Mitkä yritykset hyödyntävät teidän julkaistua avointa dataa? Miten?
6. Kuinka paljon datan avaaminen maksaa - rahassa ja resursseissa?

Osa-alue 5: Avoimen datan vaikuttavuus (taloudellinen näkökulma)

1. Määritelmää, miten ymmärrätte vaikuttavuuden?
2. Miten mittaatte / arvioitte (avoimen datan/perusrekisterien) vaikuttavuutta?
3. Miten mittaatte yleisemmin tietopalvelujen vaikuttavuutta?
4. Millä eri toimenpiteillä – oman organisaation sisällä tai yleisemmin – voitaisiin lisätä tiedon hyödyntämisen vaikuttavuutta?
5. Seurataanko organisaatiossa mitkä ovat käyttäjien näkökulmasta tärkeimmät tietovarannot? Pyritäänkö avoimen datan tarjontaa kehittämään tästä näkökulmasta? Millä tavalla?
6. Miten priorisoitte / päätätte mitä dataa avataan? Onko organisaatiossanne olemassa datan avaamiseen "priorisointijärjestystä" ja jos on, miten se on määritelty?
7. Miten organisaationa itse hyödytte datan avaamisesta?



Osa-alue 6: Sosiaalisia, kulttuurisia ja poliittisia näkökulmia

1. Millaista poliittista painetta on tuottaa tai julkaista dataa eri lailla kuin nyt?
2. Miten koette hallituksen tavoitteen ”avata kaikki julkiset tietovarannot vuoteen 2020 mennessä”? Mitä vahvuuksia, heikkouksia, mahdollisuuksia ja uhkia tähän liittyy? Mitä esteitä tai haasteita?
3. Millaisia kulttuurillisia arvoja liittyy avoimeen dataan?
4. Millaisia huolia avoimen datan julkaisuun liittyy?
5. Miten ymmärrätte koneluettavuuden?
6. Millainen lisenssi on käytössä avoimessa datassanne?

Osa-alue 7: Muita esteitä ja haasteita (kaikki näkökulmat)

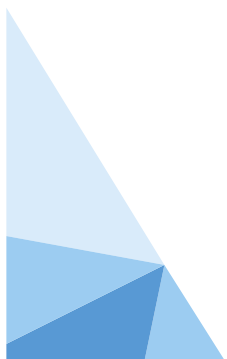
ETLA: ”Minkälaisia keräämiänne tieto(kanto)ja tai aineistoja organisaationne ei tarjoa avoimena datana ei-lakisääteisistä syistä (esim. tietosuoja)? Mistä syystä/syistä näitä tietoja ei tarjota avoimena datana ja millä edellytyksillä olisi mahdollista avata tätä dataa?”

1. Mitä käytännön esteitä organisaatiollanne on avoimen datan hyödyntämiselle ja datan avaamiselle (teknisiä, sosiaalisia, kulttuurisia, poliittisia)?
2. Millaisia lainsäädännöllisiä esteitä avoimen datan julkaisulle tai perusrekisterien avaamiselle on?
3. Avoimen datan lisäksi, mitä muita vaihtoehtoisia tapoja tiedon jakelulle ja ”tietoyhteistyölle” muiden organisaatioiden kanssa käytätte?

Osa-alue 8: Vapaa sana osio

1. Mitä asiaan liittyvää tärkeää mielestänne ei ole vielä kysytty?
2. Miten avoimen datan kustannustehokkuutta ja vaikuttavuutta parhaiten edistetään?
3. Muuta asiaa liittyvää / mieleen tulevaa ..?

Paljon kiitoksia ajastanne / vaivannäöstänne!



LIITE 4 LÄHTEET – REKISTERITIE TOJEN TARJONTA SUOMESSA

VRK

- 1) VRK: Henkilötiedot. <http://vrk.fi/henkilotiedot> Vierailtu 17.6.2016
- 2) VRK: Rakennustiedot <http://vrk.fi/rakennustiedot> Vierailtu 17.6.2016
- 3) VRK: Tietopalvelut yrityksille ja yhteisöille. <https://eevertti.vrk.fi/tietopalvelut-yrityksille-ja-yhteisoille> Vierailtu 15.6.2016
- 4) VRK: Tietopalvelut viranomaisille ja tutkimuskäyttöön. <https://eevertti.vrk.fi/tietopalvelut-julkishallinnolle> Vierailtu 15.6.2016
- 5) Valtiovarainministeriö: Perusta Nykytilan kuvaus: Liite 1. Perustietovarantojen kuvaukset, 30.12.2013. s. 5–8
- 6) Valtiovarainministeriön asetus Väestörekisterikeskuksen suoritteiden maksuista vuosina 2017 ja 2018 (1396/2016). <http://www.finlex.fi/fi/laki/alkup/2016/20161396> Vierailtu 7.2.2017
- 7) Avoindata.fi: Organisaatiot – väestörekisterikeskus <https://www.avoindata.fi/data/fi/organization/vaestorekisterikeskus> Vierailtu 2.2.2017

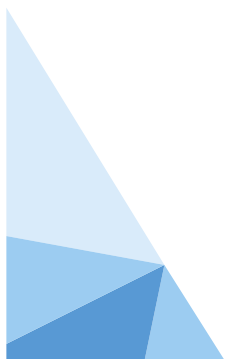
MML

- 1) MML: Tuotekuvaukset – Maastotietokanta. <http://www.maanmittauslaitos.fi/digituotteet/maastotietokanta> Vierailtu 16.6.2016
- 2) Maa- ja metsätalousministeriön asetus Maanmittauslaitoksen maksuista sekä kaupanvahvistuksesta perittävistä maksuista ja korvauksista (1386/2016). <http://www.finlex.fi/fi/laki/alkup/2016/20161386> Vierailtu 7.2.2017
- 3) MML: Tuotekuvaukset – Nimistö. <http://www.maanmittauslaitos.fi/digituotteet/nimisto> Vierailtu 16.6.2016
- 4) MML: Tuotekuvaukset – Laserkeilausaineisto. <http://www.maanmittauslaitos.fi/digituotteet/laserkeilausaineisto> Vierailtu 16.6.2016
- 5) MML: Tuotekuvaukset – Maanmittauslaitoksen ilmakekuva. <http://www.maanmittauslaitos.fi/digituotteet/maanmittauslaitoksen-ilmakekuva> Vierailtu 16.6.2016
- 6) MML: Tuotekuvaukset – Maanmittauslaitoksen ortokuva. <http://www.maanmittauslaitos.fi/digituotteet/maanmittauslaitoksen-ortokuva> Vierailtu 16.6.2016
- 7) MML: Kiinteistötietojärjestelmä – Kiinteistörekisteri. <http://www.maanmittauslaitos.fi/ammattilaisille/kiinteistotiedot/kiinteistotietojarjestelma-ktj/kiinteistorekisteri> Vierailtu 16.6.2016
- 8) MML: Aineistopalvelut – Kiinteistörekisteri. <http://www.maanmittauslaitos.fi/aineistot-palvelut/verkkopalvelut/aineistopalvelu-kiinteistorekisteri> Vierailtu 15.6.2016
- 9) MML: Kiinteistötietopalvelu – Tule käyttäjäksi. <http://www.maanmittauslaitos.fi/aineistot-palvelut/verkkopalvelut/kiinteistotietopalvelu/tule-kayttajaksi> Vierailtu 15.6.2016
- 10) MML: Kiinteistörekisteri – Kiinteistörekisteriaineiston hankinta. <http://www.maanmittauslaitos.fi/ammattilaisille/kiinteistotiedot/kiinteistotietojarjestelma-ktj/kiinteistorekisteri/kiinteistorekisteriaineiston-hankinta> Vierailtu 15.6.2016
- 11) MML: KTJ-aineistopalvelu – Hinnoittelu. <http://www.maanmittauslaitos.fi/aineistot-palvelut/verkkopalvelut/ktj-aineistopalvelu/hinnoittelu> Vierailtu 16.6.2016
- 12) MML: Kiinteistötietopalvelun oppilaitossopimus. <http://www.maanmittauslaitos.fi/lomakkeet/kiinteistotietopalvelun-oppilaitossopimus> Vierailtu 16.6.2016
- 13) MML: Aineistot ja palvelut – Kiinteistötietopalvelu. <http://www.maanmittauslaitos.fi/aineistot-ja-palvelut/palvelut/kiinteistotietopalvelu> Vierailtu 16.6.2016

- 14) MML: Verkkopalvelut – Kiinteistötietopalvelu. <http://www.maanmittauslaitos.fi/aineistot-palvelut/verkkopalvelut/kiinteistotietopalvelu> Vierailtu 16.6.2016
- 15) Valtiovarainministeriö / Anne Kauhanen-Simanainen: Perustietovarantojen maksuttomuuden toteutuminen ja vaikutukset v. 2015 mennessä (taustamuistio), 7.4.2015
- 16) Maa- ja metsätalousministeriön asetus Maanmittauslaitoksen maksuista sekä kaupanvahvistuksesta perittävistä maksuista ja korvauksista (1386/2016). <http://www.finlex.fi/fi/laki/alkup/2016/20161386> Vierailtu 7.2.2017
- 17) Kiinteistörekisterilaki (16.5.1985/392). <http://www.finlex.fi/fi/laki/ajantasa/1985/19850392> Vierailtu 7.2.2017
- 18) MML: Kiinteistötietojärjestelmä – Lainhuuto- ja kiinnitysrekisteri. <http://www.maanmittauslaitos.fi/ammattilaisille/kiinteistotiedot/kiinteistotietojarjestelma-ktj/lainhuuto-kiinnitysrekisteri> Vierailtu 16.6.2016
- 19) MML: Kiinteistörekisteri – Lainhuuto- ja kiinnitysrekisterin aineiston hankinta. <http://www.maanmittauslaitos.fi/ammattilaisille/kiinteistotiedot/kiinteistotietojarjestelma-ktj/lainhuuto-kiinnitysrekisteri/lainhuuto-kiinnitysrekisterin> Vierailtu 16.6.2016
- 20) MML: Kiinteistötiedot – Kiinteistöjen kauppahintarekisteri. <http://www.maanmittauslaitos.fi/ammattilaisille/kiinteistotiedot/kiinteistojen-kauppahintarekisteri/lisatietoa-kiinteistojen-kauppahintarekisterista> Vierailtu 16.6.2016
- 21) MML: Kiinteistöjen kauppahintarekisterin hankinta. <http://www.maanmittauslaitos.fi/ammattilaisille/kiinteistotiedot/kiinteistojen-kauppahintarekisteri/kiinteistojen-kauppahintarekisterin-hankinta> Vierailtu 16.6.2016
- 22) MML: Tuotekuvaukset – Kiinteistöjen kauppahintarekisteri. <http://www.maanmittauslaitos.fi/digituotteet/kiinteistojen-kauppahintarekisteri> Vierailtu 17.6.2016

PRH & Verohallinto

- 1) PRH: Tietopalvelut – Rekisterin tietosisältö. <https://www.prh.fi/fi/kaupparekisteri/tietopalvelut/tietosisalto.html> Vierailtu 17.6.2016
- 2) PRH: Virre-tietopalvelu – Ohjeet. <https://virre.prh.fi/novus/instructions> Vierailtu 16.6.2016
- 3) PRH: Kaupparekisterin tietopalvelut. <https://www.prh.fi/fi/kaupparekisteri/tietopalvelut.html> Vierailtu 16.6.2016
- 4) Valtiovarainministeriö / Anne Kauhanen-Simanainen: Perustietovarantojen maksuttomuuden toteutuminen ja vaikutukset v. 2015 mennessä (taustamuistio), 7.4.2015
- 5) PRH: Virre-tietopalveluhinnasto. <https://virre.prh.fi/novus/priceList> Vierailtu 6.6.2016
- 6) Työ- ja elinkeinoministeriön asetus patentti- ja rekisterihallituksen maksullisista suoritteista (758/2016). <http://www.finlex.fi/fi/laki/alkup/2016/20160758> Vierailtu 7.2.2017
- 7) PRH: Säätöt – Säätörekisterin tietopalvelut. <https://www.prh.fi/fi/saatiorekisteri/palvelut.html> Vierailtu 17.6.2016
- 8) PRH: Yhdistysnetti. <http://yhdistysrekisteri.prh.fi/etusivu.htx> Vierailtu 17.6.2016
- 9) PRH: Yhdistykset – Yhdistysrekisteri. <https://www.prh.fi/fi/yhdistysrekisteri.html> Vierailtu 17.6.2016
- 10) PRH: Avoin data – Tietoa avoimen datan rajapinnoista. <http://avoindata.prh.fi/> Vierailtu: 17.6.2016
- 11) Verohallinto: Avoin data. [https://www.vero.fi/fi-FI/Avoin_data\(29357\)](https://www.vero.fi/fi-FI/Avoin_data(29357)) Vierailtu 17.6.2016
- 12) Valtiovarainministeriön asetus Verohallinnon suoritteiden maksullisuudesta (1574/2015). <http://www.finlex.fi/fi/laki/alkup/2015/20151574> Vierailtu 29.6.2016
- 13) Valtiovarainministeriö / Anne Kauhanen-Simanainen: Perustietovarantojen maksuttomuuden toteutuminen ja vaikutukset v. 2015 mennessä (taustamuistio), 7.4.2015



Trafi

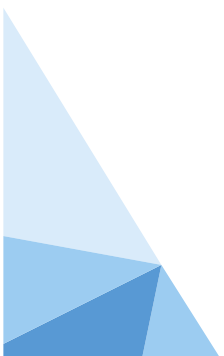
- 1) TRAFI: Tietopalvelut – Avoin data Trafissa. http://www.trafi.fi/tietopalvelut/avoin_data Vierailtu 17.6.2016
- 2) Liikenne- ja viestintäministeriön asetus Liikenteen turvallisuusviraston maksullisista suoritteista (1585/2015). <http://www.finlex.fi/fi/laki/alkup/2015/20151585> Vierailtu 29.6.2016
- 3) TRAFI: Tietopalvelut – Trafin rekisterit. http://www.trafi.fi/tietopalvelut/trafin_rekisterit Vierailtu 17.6.2016
- 4) TRAFI: Trafin rekisterit – Tietojenluovutus periaatteet. http://www.trafi.fi/tietopalvelut/trafin_rekisterit/tietojenluovutusperiaatteet Vierailtu 17.6.2016
- 5) Valtiovarainministeriö / Anne Kauhanen-Simanainen: Perustietovarantojen maksuttomuuden toteutuminen ja vaikutukset v. 2015 mennessä (taustamuistio), 7.4.2015
- 6) TRAFI: Trafin rekisterit – Palveluntarjoajat. http://www.trafi.fi/tietopalvelut/trafin_rekisterit/palveluntarjoajat Vierailtu 17.6.2016
- 7) TRAFI: Trafin rekisterit – Trafista aineistoa tutkimukseen? http://www.trafi.fi/tietopalvelut/trafin_rekisterit/tietojenluovutusperiaatteet/tietojenluovutus_tutkimuskayttoon Vierailtu 17.6.2016
- 8) Kirsi Pulkamo, Leila Hanhela-Lappeteläinen, Hannu Parkkisenniemi: Liikenteen turvallisuusviraston tietovarantojen avaaminen ja maksuttomuus (Trafin tiivistelmä Pertivalle 30.3.2015). <https://wiki.julkict.fi/julkict/juhta/juhta-n-jaostot/perustietovaranto-jaosto/pertivan-kokous-30-3.2015/tietojen-maksuttomuus-trafi-yhteenveto/view> Vierailtu 18.6.2016

Tilastokeskus

- 1) TK: Avoin data ja rajapinnat. <https://www.stat.fi/org/avoindata/index.html> Vierailtu: 17.6.2016
- 2) TK: Tuotteet ja palvelut – Px-Web-StatFin -tilastotietokantojen käyttöohje. http://tilastokeskus.fi/tup/tilastotietokannat/pxweb_ohje_2015.html Vierailtu 17.6.2016
- 3) TK: Tuotteet ja palvelut – Tilastotietokannat. <https://tilastokeskus.fi/tup/tilastotietokannat/index.html> Vierailtu: 17.6.2016
- 4) TK: Tuotteet ja palvelut – Mikroaineistot. <http://tilastokeskus.fi/tup/mikroaineistot/index.html> Vierailtu: 17.6.2016
- 5) TK: Tuotteet ja palvelut – Hinnasto. <https://tilastokeskus.fi/tup/tilastotietokannat/hinnasto.html> Vierailtu 17.6.2016
- 6) TK: Tuotteet ja palvelut – PC-Axis. <http://www.stat.fi/tup/pcaxis/pcaxis2006.html> Vierailtu 17.6.2016
- 7) TK: Valtiovarainministeriön asetus Tilastokeskuksen suoritteiden maksullisuudesta (1472/2015). <https://www.stat.fi/org/lainsaadanto/vmmaksuasetus.html> Vierailtu 29.6.2016

Ilmatieteen laitos

- 1) Ilmatieteen laitos: Avoin data – Avatut ja avattavat tietoaineistot. <http://ilmatieteenlaitos.fi/avoin-data-avattavat-aineistot> Vierailtu 21.6.2016
- 2) Ilmatieteen laitos: Ilmatieteen laitoksen avoin data. <https://ilmatieteenlaitos.fi/avoin-data> Vierailtu 21.6.2016
- 3) Liikenne- ja viestintäministeriön asetus Ilmatieteen laitoksen maksullisista suoritteista (1357/2015). <http://www.finlex.fi/fi/laki/alkup/2015/20151357> Vierailtu 29.6.2016
- 4) Ilmatieteen laitos: Kysymyksiä avoimesta datasta. <http://ilmatieteenlaitos.fi/kysymyksia-avoimesta-datasta> Vierailtu 29.6.2016



Luke

- 1) LUKE: Palvelut – Tilastopalvelu. <https://www.luke.fi/palvelut/tilastopalvelu/> Vierailtu 21.6.2016
- 2) LUKE: Tietoa tilastoista – Tiedonluovutus ruoka- ja luonnonvaratilastoista. http://stat.luke.fi/tiedonluovutus-maaseutuelinkeinotilastorekisterista_fi Vierailtu 21.6.2016
- 3) LUKE: Tilastotietoa kaikille. <https://www.luke.fi/tilastotietoa-kaikille/> Vierailtu 21.6.2016
- 4) LUKE: Tilastotietokannan käyttö. stat.luke.fi/sites/default/files/tilastotietokanta_ohje_evuosikirja.pdf Vierailtu 21.6.2016
- 5) Maa- ja metsätalousministeriön asetus Luonnonvarakeskuksen maksullisista suoritteista (1157/2014). <http://finlex.fi/fi/laki/alkup/2014/20141157> Vierailtu 29.6.2016

Liikennevirasto

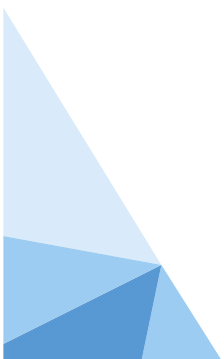
- 1) Liikennevirasto: Aineistot. <http://www.liikennevirasto.fi/aineistot> Liikennevirasto: Avoin data – Tietoaaineistot. <http://www.liikennevirasto.fi/avoindata/tietoaaineistot#.V2pK2alwP88> Vierailtu 21.6.2016
- 2) Liikennevirasto: Aineistot – Avoin data. <http://www.liikennevirasto.fi/avoindata#.V2k-1CaKc388> Vierailtu 21.6.2016
- 3) Liikennevirasto: Tilastot. <http://www.liikennevirasto.fi/tilastot#.V2pLralwP88> Vierailtu 21.6.2016
- 4) Liikennevirasto: Tilastot – Tietopalvelut ja aineistot. <http://www.liikennevirasto.fi/tilastot/henkiloliikennetutkimus/tietopalvelut-ja-aineistot#.V2k5KqKc388> Vierailtu 21.6.2016
- 5) Liikenne- ja viestintäministeriön asetus Liikenneviraston maksullisista suoritteista (1026/2014). <http://www.finlex.fi/fi/laki/alkup/2014/20141026> Vierailtu 29.6.2016

THL

- 1) THL: Avoin data. <https://www.thl.fi/fi/tilastot/tilastotietokannat/avoin-data> Vierailtu 20.2.2017
- 2) THL: Yhteistyötilat – THL:n avoimen datan tietomuodon kuvaus. <https://yhteistyotilat.fi/wiki08/display/THLKA/THL%3An+avoimen+datan+tietomuodon+kuvaus;jsessionid=3391B6EDB050071A5C45920798794B18> Vierailtu 22.6.2016
- 3) THL: Tilastot – Tutkimuskäyttö. <https://www.thl.fi/fi/tilastot/tutkimuskaytto> Vierailtu 22.6.2016
- 4) THL: Tutkimuskäyttö – Aineistopyynnöt. <https://www.thl.fi/fi/tilastot/tutkimuskaytto/aineistopyynnot> Vierailtu 22.6.2016
- 5) Sosiaali- ja terveysministeriön asetus Terveysten ja hyvinvoinnin laitoksen maksullisista suoritteista (1572/2015). <http://www.finlex.fi/fi/laki/alkup/2015/20151572> Vierailtu 29.6.2016

Kela

- 1) KELA: Tilastot – Tilastotietokanta Kelasto. <http://www.kela.fi/kelasto> Vierailtu 22.6.2016
- 2) Valtiovarainministeriö / Anne Kauhanen-Simanainen: Perustietovarantojen maksuttomuuden toteutuminen ja vaikutukset v. 2015 mennessä (taustamuistio), 7.4.2015





VALTIONEUVOSTON
SELVITYS- JA TUTKIMUSTOIMINTA

tietokayttoon.fi

ISSN 2342-6799 (pdf)

ISBN 978-952-287-387-3 (pdf)

