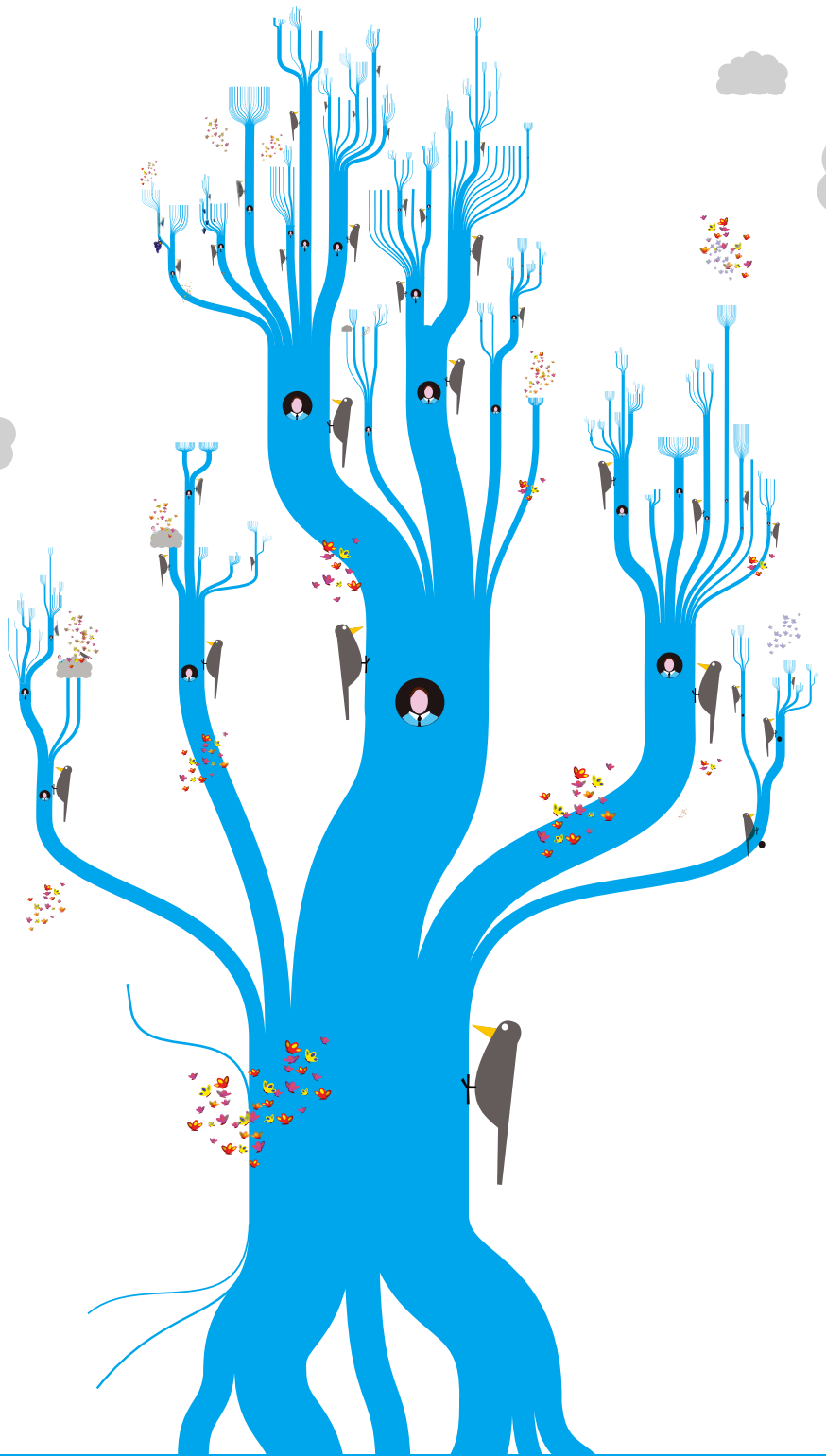




Liikenne- ja
viestintäministeriö



Antti Poikola
Petri Kola
Kari A. Hintikka

Julkinen data^{cc}

johdatus tietovarantojen avaamiseen

Case: Veropuu

Kannen kuva: Peter Tattersall – Veropuu

Peter Tattersall voitti Suomen Kansalaisosallistujan työkalut -kilpailun (Apps for Democracy Finland) idea-kategorian 2009. Peterin idea on oikeastaan aika yksinkertainen. Suomen budjetti on työlästä luettavaa jopa ammattilaiselle. Budjettiaineisto sellaisenaan on ollut julkisesti saatavilla jo pitkään, ja sitä on pyritty visualisoimaan myös median toimesta.

Veropuu on ehdotus www-palveluksi, joka visualisoi valtion, kunnan, yksittäisen julkishallinnon organisaation tai vaikka yhtiön tulot ja menot. Puun juurina ovat tulot, jotka liittyvät runkoon ja edelleen haarautuvat menoina oksiin. Lehdet ja hedelmät ovat hyötyjä, joita käytetyllä rahalla saavutetaan. Juurien ja oksien paksuudet esittävät eri tulonlähteiden ja menokohteiden suuruutta.

Veropuu ja vastaavat sovellukset tarvitsevat avointa dataa koneluettavassa muodossa. Valtiokonttorin Netra-palvelu (www.netra.fi) tarjoaa informaatiota valtion toiminnasta, resursseista ja tuloksellisuudesta, mutta palvelu on suunniteltu erityisesti valtioneuvoston tarpeisiin. Mitä Suomessa pitäisi tehdä, että veropuun toteuttaminen olisi mahdollisimman helppoa? Se on kysymys, jonka ympärille tämä opas rakentuu.

Avoimesti lisensoitu opas

Tämä teos on lisensoitu Creative Commons Nimeä 1.0 Suomi lisenssillä. Teoksen uudelleenkäytön yhteydessä pitää mainita sekä kirjoittajat (Antti Poikola, Petri Kola ja Kari A. Hintikka) että rahoittaja (Liikenne- ja viestintäministeriö). Nähdäksesi lisenssin vieraile <http://creativecommons.org/licenses/by/1.0/fi/> sivulla.

Valittu lisenssi tarkoittaa, että oppaan sisältö on vapaasti käytettävissä, kunhan alkuperäislähteeseen viitataan. Ilman Creative Commons -lisenssiä kirjallisen julkaisun tekijänoikeudet rajoittaisivat sen käyttömahdollisuuksia datan avaamisen tukemiseksi. Oppaassa kannustetaan vastaavien sallivien lisensointikäytäntöjen laajempaa käyttöönottoa hallinnossa.

Ulkoasu: **Kirmo Kivelä** ja **Inari Savola**

ISBN 978-952-243-145-5 (painotuote)

ISBN 978-952-243-146-2 (verkkojulkaisu <http://www.lvm.fi>)

Edita Prima Oy, Helsinki 2010



Julkinen data^{cc}

johdatus tietovarantojen avaamiseen

Antti Poikola | Petri Kola | Kari A. Hintikka

Opas on tehty lyhyessä ajassa tulevia kehityskulkuja ennakoiden ja suuntautuen suomalaisen tietoyhteiskunnan sekä EU:n kehitysnäkymiin.

Oppaan kirjoittaminen tukeutuu laajaa haastatteluaineistoon, josta monet oppaassa esitetyt ajatukset ovat peräisin. Aineiston kerääminen aloitettiin keväällä 2009 Suomen Akatemian rahoittamassa Somus-hankkeessa ja sitä on täydennetty vuoden 2010 alussa vartavasten opasta varten tehdyllä haastattelu-kierroksella.

On ollut ilo tutustua haastateltaviin ja nähdä kuinka tietoyhteiskuntaa eri puolilla Suomea rakennetaan. Kiitos: Sari Aalto-Matturi (Oikeusministeriö), Reijo Aarnio (Tietosuojavaltuutettu), Olli Ahonen (Valtiokonttori), Henni Ahvenlampi (Kulttuurikeskus Cajsa), Ulla Ala-Ketola (www.ymparisto.fi), Päivi Antikainen (Liikenne- ja viestintäministeriö), Jussi Arpalahti (Tilastokeskus), Aki Asola (Eduskunta), Pirkka Aunola (YLE), Hans Baumgartner (Tilastokeskus), Yrjö Benson (Valtion IT-johtaja), Jyri Engeström (yrittäjä, sijoittaja), Liisa Ero (Liikenne- ja viestintäministeriö), Esa Hirvonen (www.ymparisto.fi), Elina Huhta (Kulttuurikeskus Cajsa), Risto Joki (Ulkoministeriö), Jan Juslen (Liikennevirasto), Jyrki Koskinen (IBM), Päivi Kuusisto (Tampereen Kaupunki), Marko Latvanen (Suomi.fi), Harri Lehtinen (Tilastokeskus), Raija Leppäjarvi (Ulkoministeriö), Antti Leskinen (Tampereen Kaupunki), Asta Manninen (Helsingin tietokeskus), Heli Mikkela (Tilastokeskus), Olli Mustajarvi (Eduskunta), Rolf Nordqvist (Bisnode), Petteri Nyman (Eduskunta), Leila Oravisto (Helsingin Kaupunki), Kristiina Pietikäinen (Liikenne- ja viestintäministeriö), Pirkko Romakkaniemi (Suomi.fi), Antti Rainio (Maanmittauslaitos), Oili Salminen (Oikeusministeriö), Päivi Tahvanainen (www.ymparisto.fi), Ari Tammi (Helsingin Kaupunki) ja Kalle Ylä-Anttila (YLE).

Kiitos kaikille, joilta olemme saaneet asiantuntevaa apua oppaan kirjoittamisessa. Erityisesti haluamme mainita seuraavat henkilöt: Henri Bergius (Nemein), Antti Eskola (Työ- ja elinkeinoministeriö), Tuula Haavisto (Tampereen Kaupunki), Auli Harju (Somus), Jari Honkonen (HSL), Ossi Kuittinen (Sitra), Ville Meloni (Forum Virium), Ville Oksanen (EFFI), Ilkka Pirttimaa, Teemu Ropponen (Somus), Samuel Rinnetmäki, Petri Tonteri (Tampereen Kaupunki) ja Kim Viljanen (SeCo).

Kiitos kaikille Kansalaisosallistujan työkalut (Apps for Democracy Finland) -kilpailuun osallistuneille. Ilman teidän panostanne, olisi hyvin vaikea konkreettisesti ymmärtää mitä avoin data voi Suomessa merkitä. Oppaan lopussa (liite 3) on lyhyesti esitelty kilpailun osallistujat projekteineen.

Lopuksi vielä kiitos Liikenne- ja viestintäministeriön Taru Rastaalle.

[Julkishallinnolla on tietoa](#), jonka avaaminen ja maksuton käyttö on saanut tukea yhteiskunnallisessa kehityskeskustelussa ja kansainvälisistä esimerkeistä.

Arjen tietoyhteiskunnan neuvottelukunta on tuonut esille julkisen tiedon saatavuuden merkitystä niin tuottavuuden, kilpailukyvyn kuin hyvinvoinnin kannalta. Viestinnän elinkeinopolitiikan työryhmä on esittänyt, että tietojen tulisi olla helpommin yritysten käytettävissä digitaalisissa verkkopalveluissa. Selvitysmies Paul Paukun raportissa ”Julkiset sähköiset palvelut – kehittämisen pullonkaulat” esitetään visio Suomesta edelläkävijänä avoimen tiedon yhteiskuntana. Uuden kansallisen tietoyhteiskuntastrategian yhdeksi teemaksi on nostettu tiedon rooli perusinfrastruktuurina. Parhaillaan ministeriöiden asettamat työryhmät etsivät ratkaisuja julkisen tiedon saatavuuden helpottamiseen. Myös Elinkeinoelämän valtuuskunta on peräänkuuluttanut toimia avoimen tiedon puolesta. EU:n komissio on todennut, että jäsenvaltioiden on tehostettava toimia, joilla julkisen datan tarjoama potentiaali voitaisiin hyödyntää.

Toimintaan sovellettavaa tietämystä datan avaamiseksi on kuitenkin vähän. Aiheesta ei tätä opasta aiemmin ole ollut kokoavaa tietoa suomeksi. Tämä opas pyrkii kertomaan, miten datavarantoja voidaan avata hallitusti. Tietoaineistojen ympärille syntyy palveluita ja tietoa jalostavia yhteisöjä. Tämän ekosysteemin rakennuspalikoiden esittely on kirjan keskeistä antia. Opas tarjoaa myös suunntaa avoimen datapolitiikan periaatteille, joista meidän tulee kansallisesti sopia.

Tietoyhteiskunnassamme on laajasti herätty näkemään tiedon avoimuus ratkaisuna yhteiskunnallisiin ja taloudellisiin haasteisiin. Uusien liiketoimintamahdollisuuksien lisäksi julkisen tiedon avoimuus lisää hallinnon tehokkuutta. Edut eivät rajoitu vain verkkopalveluihin, vaan vaikutukset liittyvät ympäristöömme laajasti esimerkiksi liikkumisen ratkaisuissa tai aluesuunnittelussa. Parhaimmillaan avoin data luo yhdessä tekemisen toimintakulttuuria, johon Internetin yhteisöllinen ja teknologinen kehitys tarjoaa tilaisuuden. Tässä muutoksessa lainsäädäntö ja hallinnolliset käytänteet tulee saattaa ajan tasalle.

Avoimen datan vaikutusten arviointi on toistaiseksi suuntaa-antavaa. Emme kuitenkaan voi jäädä odottamaan. Tarvitsemme ennakkoluulotonta kokeilua, sillä Suomella sanotaan olevan laajojen ja laadukkaiden julkisten tietovarantojen sekä teknisen osaamisen ansiosta erityinen asema nousta johtajaksi avoimen tiedon yhteiskuntana.

Kirjoittajakolmikον kyky jalostaa kotimaista ja kansainvälistä asiantuntemusta on vakuuttavasti täyttänyt työn tarkoituksen kertoa, miten tietovarantoja käytetään innovatiivisesti ja tuottavasti. Nyt oppaan tehtävä on kannustaa tiedon tarjoajia ja käyttäjiä luomaan näitä käytäntöjä.

Helsingissä 15.3.2010
Liikenne- ja viestintäministeriö
Taru Rastas

Julkishallinnon tuottamat ja omistamat tietovarannot ovat herättäneet viime aikoina runsaasti keskustelua Suomessa ja Euroopan unionissa. Vuoden 2010 alkupuoliskolla Suomessa on perustettu työryhmiä ja tuotu julki useita kannanottoja nykyisen toimintatavan ja lainsäädännön tarkistamiseksi.

Suomen nykylinjaus julkishallinnon tuottamasta datasta perustuu vuonna 1992 hyväksyttyyn maksuperustelakiin, jonka mukaan monissa tapauksissa julkishallinnon tuottaman datan käytöstä peritään korvaus. Maksuperustelaki luotiin aikana ennen internetiä. Nykyään, internetin myötä, datan tarjoamisen kustannukset ovat oleellisesti vähäisempiä kuin tuolloin.

Julkishallinnon datan maksuton jakaminen hyödyttäisi kokonaisedullisesti suomalaista liike-elämää, kansalaistoimintaa ja hallinnon tehostamista. Aiheesta ei ole vielä tarkkoja laskelmia, mutta tutkimusten ja raporttien mukaan nykyisessä mallissa suuri osa datanluovutuksista kertyneistä tuloista saadaan hallinnolta itseltään. Oppaan esittämässä näkökulmassa hallinnon datan avaaminen maksuttomasti olisi Suomelle kokonaisedullisempaa kuin nykykäytäntö.

Opas kuvaa käytännönläheisesti ja yleisellä tasolla prosessin hallinnon tietovarantojen avaamiseksi yleiseen maksuttomaan käyttöön. Yhdysvalloissa sekä Isossa-Britanniassa hallinnon datan avaaminen on jo otettu osaksi toimintapolitiikkaa ja strategioita. Opas hahmottaa hallinnon datan avaamista myös laajassa yhteiskunnallisessa kehityksessä.

Datan avaaminen alkaa organisaation omien aineistojen evaluoinnilla. Organisaation koon ja luonteen vaihdellessa tämä voi olla laajamittainenkin prosessi, mutta toisaalta kaikkea ei tarvitse tehdä kerralla. Avaamisessa voidaan edetä vaiheittain helpoista aineistoista, matkan varrella oppien, kohti vaikeampia. Datainventaarion myötä voi löytyä yksittäisen osaston tuottamaa dataa, josta organisaatio ei ole edes tietoinen tai jota ei ole osattu hyödyntää.

Inventaarion myötä organisaatio pystyy luomaan oman strategiansa ja tavoitteensa datansa hyödyntämiseksi. Hyötyjä voivat olla esimerkiksi uudet käyttötarkoitukset, yhteistyökumppanit tai organisaation roolin jalostuminen. Opas tarjoaa työkalupakin datan avoimuuden arvioimiseksi. Inventaarion jälkeen data olisi hyvä saattaa koneluettavaan muotoon. Yhä useammin dataa käytetään internet- ja mobiilisovelluksissa, joiden avulla muodostetaan käyttäjälle lisäarvoa sen sijaan, että ihminen hakisi selaimella itse vastaavan datan internetistä. Suomi on tuottanut erittäin laadukkaita datavarantoja, mutta usein aineistot julkaistaan vain esimerkiksi pdf-muodossa, jota ei voida helposti ohjelmallisesti jatkojalostaa.

Julkiseen dataan ja sen julkistamiseen liittyy useita lakeja, direktiivejä ja suosituksia, jotka on hyvä huomioida avaamisprosessissa. Näitä ovat muun muassa julkisuuslaki, tietosuojalainsäädäntö, maksuperustelaki, tekijänoikeuslaki, kansainväliset suositukset, kilpailulainsäädäntö sekä EU-direktiivit. Lähtökohtaisesti mikään em. laeista ja säädöksistä ei estä datan avaamista, mutta niissä on kohtia, jotka on hyvä tuntea jotta avaaminen tapahtuu hallitusti.

Datan avaamista voi ajatella vuorovaikutteisena prosessina siksi, että usein parhaimmat käyttötarkoitukset ja käyttötavat syntyvät organisaation ulkopuolella. Opas rinnastaa datan avaamisen ekosysteemiksi, jossa erityyppiset toimijat tarjoavat dataa ja sitä hyödynnetään ilman vastavuoroisuutta siten, että kokonaisuus hyödyttää kaikkia osapuolia. Toinen tapa jäsentää avointa dataa on infrastruktuurinen ajattelu. Avoimen datan voi mieltää mahdollistajaksi sekä sisällöksi, joka on osa teknistä infrastruktuuria katujen tai sähkön tavoin.

Oppaan lopussa esitetään joukko ehdotuksia suomalaisten tietovarantojen maksuttomalle avaamiselle. Ehdotukset jakautuvat viiteen aihealueeseen:

- 1)** Avoimen datan perusinfrastruktuurin synnyttäminen
- 2)** Lainsäädännön muutos ja kansallinen ohjaus
- 3)** Hallinnon sisäiset käytännöt tukemaan avointa dataa
- 4)** Julkishallinnon avoimen datan hyötyjen jalkauttaminen
- 5)** Suomen kansainvälisen edelläkävijyyden ja tahtotilan mahdollistaminen

Suomessa ei vielä ole datan avaamista koordinoivia toimielimiä, kuten esimerkiksi Isossa-Britanniassa. Jotta eteneminen olisi koordinoitua ja yksittäisen organisaation kannalta vaivatonta, Suomeen voitaisiin luoda hallinnon avoimen datan selvitystoimisto (clearing house). Se koordinoisi käytännön kysymyksiä, opastaisi hallintoa ja toimisi myös mahdollisten ongelmatilanteiden ratkaisijana esimerkiksi Kuluttajaviraston tavoin. Niin ikään Suomelle voitaisiin kehittää datakatalogi, jossa olisi keskitetysti tarjolla tiedot kaikista hallinnon tuottamista julkisista aineistoista.

Esipuhe	5
Tiivistelmä	7
Johdanto	11
Avoin data monialaisen hyödyn lähteenä	11
Oppaan sisältö	12
Data käsitteenä	13
Data ekosysteemin osana	15
Tahtotila: Julkinen data avoimesti saataville ja hyödynnettäväksi	16
Tavoite 1: Uutta tietoa tuottavat koulutetut suomalaiset	16
Tavoite 2: Talkoot ja yhteistoiminnan perinne	17
Tavoite 3: Laadukkaat tietovarantomme kansainvälisen profiloitumisen perustaksi	18
Tavoite 4: Dataa myös koneille	18
Tavoite 5: Pienestä liikkeelle	19
1. Elävä palvelu	21
1.1 Mihin avointa dataa voidaan käyttää?	22
1.2 Käyttäjälähtöinen innovointi	25
1.3 Ketterät menetelmät datan avaamisessa	26
1.4 Organisaatio mahdollistajana	26
1.4.1 Avoin toiminta ja viestintä datan hyödyntäjille	27
1.4.2 Innovaatiokilpailu kannustimena toimintaan	28
2. Organisaation näkökulmia avoimuuteen	31
2.1 Datavarantoja toimijoita ja rooleja	32
2.2 Datan avoimuuden mittareita	34
2.2.2 Kokonaisuus	35
2.2.3 Käyttöehtojen tasa-arvoisuus	35
2.2.4 Alkuperäisyys ja ajantasaisuus	36
2.2.5 Laillinen ja vapaa uudelleenkäytettävyys	36
2.2.6 Maksuttomuus	36
2.2.7 Koneluettavuus	37
2.2.8 Formaatin avoimuus	37
2.2.9 Ymmärrettävyys	38
2.3 Datavarantojen inventaario- ja julkaisuprosessi	38
2.3.1 Datasta ilmoittaminen	40
2.3.2 Pilottiaineistojen julkaiseminen	41
2.3.3 Dokumentointi ja käyttötilanteet	42
2.3.4 Tietoarkkitehtuurin kehittäminen ja sanastotyö	43
3. Lupa julkaista ja käyttää	45
3.1 Datan julkisuuteen vaikuttava lainsäädäntö	47
3.2 Datan maksullisuuteen vaikuttava lainsäädäntö ja sääntely	48
3.3 Datan käyttöoikeuksiin vaikuttava lainsäädäntö	48
3.4 Kilpailulainsäädäntö	49
3.5 Julkishallinnon dataan liittyvät EU-direktiivit	50
3.5.1 Julkisen datan kaupallinen hyödyntäminen ja PSI-direktiivi	50
3.5.2 INSPIRE-direktiivi ja paikkatietolaki	51

4. Avoimen datan taloudellisia näkökohtia	53
4.1 Muuttunut toimintaympäristö	55
4.2 Maksuperustelaki ajalta ennen internetiä	56
4.3 Irtiottokustannus EU:n maksimisuosituksena	56
4.3.1 Hinnoittelun vaikutus datan uudelleenkäyttöön	57
4.4 Data julkishyödykkeenä	58
4.5 Aikakustannus	59
4.6 Iso-Britannian innovaatiovaluutta	60
4.7 Maksuttomuus kokonaisedullisuutena	61
5. Tekninen valmistautuminen	63
5.1 Suunnittelukysymyksiä	65
5.2.1 Rajapinnat ja esitysmuodot datan mukaisesti	66
5.2.2 Koneluettavat lisenssit	67
5.3 Web-arkkitehtuuri	67
5.4 Sisällön esitysmuotoja	69
5.4.1 Ilmoitukset muutoksista syötteiden avulla	69
5.4.2 Ajantasainen verkko	69
5.4.3 Paikkatieto uudelleenkäytettävässä muodossa	70
5.4.4 Asiakirjojen julkaiseminen	71
5.5 Rajapinnat, sovellukset ja palvelut	71
5.6 Linkitetty data	72
6. Avoimen datan infrastruktuuri	75
6.1 Datakatalogi – kaikki julkinen data yhdeltä luukulta	76
6.1.1 Datakatalogien tekninen yhteensopivuus	77
6.2 Taustatyö datakatalogin takana	77
6.2.1 Julkisen datan avaamispalvelu	78
6.3 Suomalaisia hankkeita	78
7. Ehdotuksia toimenpiteiksi	81
7.1 Avoimen datan perusinfrastruktuurin synnyttäminen	82
7.2 Lainsäädännön muutos ja kansallinen ohjaus	83
7.3 Hallinnon sisäiset käytännöt tukemaan avointa dataa	84
7.4 Julkishallinnon avoimen datan hyötyjen jalkauttaminen	85
7.5 Suomen kansainvälisen edelläkävijyyden ja tahtotilan mahdollistaminen	85
Liitteet	87
Lähdeluettelo	91

Hallinto tuottaa jatkuvasti suuren määrän dataa, joka voisi olla tehokkaamminkin käytössä huomioiden tietovarantojemme laadukkuus.

Julkinen hallinto omistaa ja hallinnoi laajoja, rahallisesti ja yhteiskunnallisesti arvokkaita informaatiovarantoja. Suurin osa näistä raakamuotoisena datana olevista varannoista on tällä hetkellä vain niiden käytössä, joilla on pääsy aineistoihin. Datan jatkojalostukseen päätyy arvioiden mukaan vain pieni osa käytettävistä olevasta. Internetin kehitys niin yhteisöllisiltä kuin teknologisiltä ominaisuuksiltaan antaa uusia mahdollisuuksia avoimeen datapolitiikkaan. Kansallisten tietovarantojen maksuttoman käytön salliminen ja sitä tukeva toimintakulttuurin muutos ovat merkittävimpiä asioita, joita suomalaisen tietoyhteiskunnan kehittämiseksi voitaisiin tällä hetkellä toteuttaa. Tämä toisi uusia mahdollisuuksia vastata ajankohtaisiin haasteisiin. Maksuton julkinen data olisi myös suomalaisille instituutioille mahdollisuus uudistua ja oppia verkostoaikauden yhteistoimintatapoja sekä suunnata suomalaisen yhteiskunnan kehittämistä uudenaikaiseen yhdessä tekemisen kulttuuriin.

Viime aikoina ajatukset sallia julkishallinnon datan käyttäjiksi niin yksityiset yritykset, tutkimuslaitokset kuin kaikki muutkin kiinnostuneet tahot on saanut yksimielistä tukea. Keskustelu aihepiiristä on lisääntynyt meillä omien kansallisten tavoitteidemme, mutta myös kansainvälisten esimerkkien innoittamana.

Julkisen datan maksuttoman saatavuuden puolesta ovat viime aikoina esiintyneet monet tahot (ks. LVM 2009, Paukku 2009, Turkki 2009). Myös Arjen tietoyhteiskunnan neuvottelukunta on ottanut kantaa sen puolesta, että julkinen data saatettaisiin avoimesti ja laajasti kansalaisten ja elinkeinoelämän käytettäväksi (Neuvottelukunta 2009). Helmikuussa 2010 liikenne- ja viestintäministeriö ja valtiovarainministeriö asettivat kolme työryhmää selvittämään, miten julkisen informaation saatavuutta parannettaisiin ja kuinka tiedon käyttöä julkisen ja yksityisen sektorin kesken voitaisiin edistää (LVM & VM 2010).

Tätä kirjoitettaessa ajankohtaisin kannanotto avoimen datan puolesta on valtiovarainministeriön työryhmältä, joka on pohtinut tilastotoimen tehostamista ja alueellistamista. Työryhmän esittää useita toimenpiteitä, joiden tavoitteena on helpottaa julkisin varoin luotujen tietovarantojen hyödyntämistä. Parantamisalueina työryhmä nostaa esiin kuntatilastot sekä ympäristö- ja luonnonvaratilastot (VM 2010).

Enää ei nähdäksemme ole tarvetta periaatekeskusteluun avoimesta datasta, vaan erityistä painetta kohdistuu jo tietotaidon lisäämiseen ja käytäntöjen luomiseen orientoivalla opastuksella.

Avoin data monialaisen hyödyn lähteenä

Suomessa on hallinnon avoimeen ja maksuttomaan dataan liittyvässä keskustelussa tähän asti viitattu ensisijaisesti liike-elämän hyötyihin ja kansalliseen kilpailukykyyn. Markkinatalouden kärkimaassa Yhdysvalloissa julkista dataa koskeva lainsäädäntö on jo pitkään perustunut ajatukselle, että verovaroin tuotetun datan tulee olla maksutta kaikkien saatavilla. Tämä lähtökohta on johtanut merkittävään julkisen datan uudelleenkäyttöön ja taloudelliseen toimeliaisuuteen. Suomessa, kuten muuallakin Euroopassa, lähtökohta on ollut varsin erilainen ja suhtautuminen datavarantoihin perinteisesti protektionistista. Edellytykset uudenaikaiselle ajattelulle kasvavat kuitenkin jatkuvasti, kun avoimet toimintamallit kehittyvä ainakin osaksi internetin ansiosta.

Viranomaiset keräävät dataa ensisijaisesti omaa käyttöään varten. Arvokkaat uudelleenkäytettävät datavarannot syntyvät muun toiminnan sivutuotteena. Useissa tässä oppaassa esitellyistä esimerkeistä parhaat käytötavat ovat muotoutuneet tuottajaorganisaation ulkopuolella ennalta arvaamattomissakin käyttöympäristöissä ja toisiin datalähteisiin yhdistettyinä. Monet näistä esimerkeistä ovat yllätyksellisiä niin toimijoiden kuin yhteistyökumppaneidenkin osalta.

Julkisten datavarantojen avaaminen edistää samanaikaisesti ainakin kolmea eri tavoitetta: demokratian ja hallinnon läpinäkyvyyden lisäämistä, uusien markkinoiden ja innovaatioiden syntymistä ja hallinnon sisäisen tehokkuuden lisäämistä. Näitä alueita esittelemme oppan eri luvuissa tarkemmin asiayhteyksiin sitoen. Taloudellisten hyötyjen ohella uudenlainen toimintapolitiikka vahvistaisi osaltaan myös identiteettiämme, ts. antaisi pontta pohdintoihin siitä, miten me tulevaisuudessa elämme nettiaikakaudella ja mille asioille kansallisesti tuottamamme informaatio vastaisuudessa rakentuu ja miten hyödynnämme yhdessä tuottamaamme dataa?

Avoin data ja datan koneluettavuus virtaviivaistavat prosesseja, mutta tätä mahdollisuutta ei ole vielä käytetty kovinkaan laajasti. Liikenne- ja viestintäministerin asettama selvitysmies Paul Paukku kritisoi hallintoa kyvyttömyydestä yhteistyöhön liike-elämän kanssa: “Not invented here- syndrooma: valtio rakentaa tyhjästä sitä mitä markkina jo käyttää” (2009). Avoin informaationvälitys ja läpinäkyvyys mahdollistaisivat laajemman näkemyksen, jolla voidaan edistää yhteistyötä. Julkisen datan maksuttomalla jakamisella voidaan välttää myös tahatonta osaoptimointia eli sitä, että virastot säästääkseen omista kustannuksista eivät esimerkiksi osta palveluna väestötietojärjestelmän tarjonnasta, vaan ylläpitävät itse tietokantaa omista asiakkaistaan.

Oppaan sisältö

Toistaiseksi maassamme ei ole ollut avoimen julkisen datan hyödyntämiseen opastusta tai käytäntöjä. Tämä opas pyrkii poistamaan tätä puutetta. Se on käytännönläheinen johdatus ensisijaisesti julkishallinnon työntekijöille, jotka haluavat edistää datan avaamista omassa organisaatiossaan, mutta on käyttökelpoinen myös kaikille muille aihepiiristä kiinnostuneille.

Avoin julkinen data on aiheena ja ilmiönä haastava tarkastelukohde. Siksi siitä on sekä mielenkiintoista että vaikeaa tehdä opastusta. Tässä oppaassa olemme päätyneet seuraavanlaiseen järjestykseen. Alkuun esittelemme aihepiiriä ja siihen liittyviä käsitteitä, toimintamalleja ja ominaisuuksia sekä luomme ideaalitalannekuvaa siitä, mitä avoimen datan hyödyntäminen parhaimmillaan voisi tuottaa maallemme. Seuraavaksi tarkastelu siirtyy enemmän konkreettiselle tasolle ja pohdimme avoimeen datan jakamiseen liittyviä lainsäädännöllisiä, taloudellisia ja teknisiä kysymyksiä. Lopuksi esitämme suosituksia, joita noudattamalla keskustelu ja toiminta avoimen datan alueella etenisi ja tulisi myös käytännöksi.

Yksi oppaan perusteemoista käsittelee julkishallinnon tuottaman datavarannon avaamista maksuttomasti kaikkien käytettäväksi niin, että dataa voidaan jalostaa koneluettavassa muodossa. Opas esittelee työkalujen ja runsaiden tapausesimerkkien kautta mitä kaikkea on hyvä huomioida omia datavarantoja avattaessa. Käsittelemme myös hallinnon ja käyttäjäyhteisöjen välistä yhteistyötä, jolla voidaan tukea datan uudelleenkäyttöä ja näin lisätä sen yhteiskunnallista ja kansantaloudellista vaikuttavuutta.

Oppaaseen on myös tiivistetty keskeisimmät kansainväliset opit suomalaisen toimintaympäristöön ja kulttuuriin sovitettuina ajatuksella, että osaa niistä voitaisiin soveltaa heti käytännössä. Yksityiskohtaista opastusta käytännön sovelluksiin on kuitenkin vaikea antaa, sillä hyvät käytännöt ovat monilta osin vasta muotoutumassa. Opas toivottavasti viitoittaa tietä ensimmäisiin kokeilui-

hin, käytäntöjen kehittämiseen ja toimii keskusteluun virittäjänä.

Oppaan sisällössä on nostettu esille niin ikään keskeisiä kysymyksiä datan teknisestä avaamisesta, käyttöönoton arvioinnista ja yhteisöjen kanssa työskentelemistä. Aihepiirin laajuuden vuoksi käsittelemme yksittäisiä asioita jokseenkin yleisellä tasolla, mutta opas tarjoaa asioista kiinnostuneille runsaasti lähteitä teemojen syventämiseen. Olemme rajanneet käsittelystä monia avoimesta datasta kumpuvia erittäin kiinnostavia teemoja, kuten tietoturvan, hallinnon it-arkkitehtuurin, digitaalisen kuilun, rekisterien käytön, dokumentinhallinnan ja yleisen tekijänoikeus -keskustelun.

Johdanto ja ensimmäinen luku antavat taustaa aiheeseen yleisesti tutustuvalla lukijalla. Johdanto esittelee oppaassa käyttämäämme termistöä. Lisäksi johdannossa esitämme lukijan orientoimiseksi viisi tavoitetta Suomen tahtotilaksi ja hallinnon tietovarantojen avaamiseksi lähitulevaisuudessa. Näitä tavoitteita toivon mukaan kansantajuistavat niissä esiteltyt suomalaiset innovaatiot. Tavoitteita ei välttämättä tarvitse lukea, mikäli on enemmän kiinnostunut varsinaisesta opastuksesta. Tavoitteet johdattelevat avoimen datan laajempiin viitekehyksiin.

Ensimmäinen luku: *Elävä palvelu* jäsentää organisaation suhtautumista yhteistyöhön, verkostoihin sekä muuhun oman organisaation ulkopuoliseen toimintaan, kun tavoitteena on mahdollistaa julkishallinnon dataan pohjautuvien uusien syntyminen.

Luvut 2, 3, 4 ja 5 avustavat organisaatiota muodostamaan omaa näkökulmaa avoimeen dataan. Lukujen sisältö on koottu tapausesimerkkien sekä haastattelujen kautta. **Luku 2:** *Organisaation näkökulmia* avoimuuteen esittää viitteellisen prosessin organisaation omien tietovarantojen inventointiin ja avaamiseen, sekä työkaluja aineistojen avoimuuden arviointiin eri näkökulmista (käyttöehdot, maksuttomuus, koneluettavuus, löydettävyyys jne.). Teemoja syvennetään juridikan ja tekniikan osalta seuraavissa luvuissa. **Luku 3:** *Lupa julkaista ja uudelleenkäyttää* esittelee yhteenvetomaisesti joukon keskeisimpiä lakeja sekä asetuksia, joita on hyvä huomioida. **Luku 4:** *Avoimen datan taloudesta* kokoaa viime vuosien keskustelua siitä, miten hallinnon datan tuotantoa ja jakelua rahoitetaan nykyään, miksi siihen on päädytty Suomessa ja miten sitä voitaisiin rahoittaa. **Luku 5:** *Tekninen valmistautuminen* tuo esiin avoimen datan keskeisen ulottuvuuden; sen saattamisen tarjolle automatisoidusti ja ohjelmallisesti käsiteltävässä muodossa.

Luku 6: *Avoimen datan infrastuktuuuri* esittelee kansallista avoimen datan infrastruktuuria, kuten datakatalogeja ja poikkihallinnollisia toimijoita, jotka auttavat muita toimijoita datavarantojen avaamisessa ja hyödyntämisessä.

Luku 7: *Suosituksia Suomelle* esittää laajempia toimenpidekokonaisuuksia, joilla voitaisiin tukea kansallista avoimen datan infrastruktuurin luomista ja käytäntöjä. Kehitys lähtee yksilöistä ja organisaatioista, mutta kaikkea yksittäiset toimijat tai organisaatiot eivät voi järkevästi edistää, vaan tarvittaisiin kansallisen tason linjanvetoja. Oppaan loppuun olemme koonneet hyväksihavaittuja liitteitä ja työkaluja.

Data käsitteenä

Suomen kielessä termien data, informaatio (information) ja tieto (knowledge) arkikäytössä ei ole useinkaan selvää eroa. Kotimaisessa puheessa tiedolla viitataan yleensä ja tarkemmin erittelemättä niin dataan, informaatioon kuin tietoonkin. Toisaalta tiedolla on suomenkielessä - muista kielistä poiketen - myös arvollinen merkitys. Esimerkiksi 'tietoyhteiskunta' sisältää enemmän sivistykseen ja kulttuuriin liittyviä latauksia kuin muissa kielissä neutraalimpi 'informaatioyhteiskunta'. Ihminen voi olla tiedoton monessa mielessä. Datan informaation ja tiedon ohella asiaan liittyviä käsitteitä ovat sisältö ja teos, joihin liittyvät muun muassa tekijänoikeudet.

Informaatiotutkimuksen piirissä datalla tarkoitetaan merkeistä ja symboleista koostuvaa potentiaalista informaatiota. Data on raaka-ainetta, josta voidaan jalostaa merkityksellisempää informaatiota, jota puolestaan voidaan välittää edelleen (Vakkari 1999). Keskeisessä roolissa on informaation tulkinta. Informaatio muodostuu tiedoksi kun vastaanottaja tulkitsee saamansa informaation ja antaa sille merkityksen siten, että se muuttaa tulkitseen käsityksiä jostakin asiasta. Informaatiotutkimuksessa puhutaan yleisesti tiedon arvoketjusta (value chain of information), jossa irrallinen informaatio jalostuu ihmiselle käytännössä hyödylliseksi.

Kaavio. Tiedon arvoketju (Haasio, Savolainen 2004).

DATA -> INFORMAATIO -> TIETO -> TIETÄMYS -> VIISAUS

Tieto- ja informaatiokäsitteet voidaan erottaa toisistaan. Niiniluodon (2000) mukaan informaatio on yläkäsite ja tieto sen alakäsitteenä tarkoittaa tarkempaa perusteltavuuteen ja totuudenmukaisuuteen liittyvää vaatimusta. Karvosen (2000) mukaan arkikielessä tieto ja informaatio ovat melkein sama asia, mutta tieteellisessä käytössä niiden välillä on suuri ero. Informaatiosta voidaan hänen mukaansa erottaa ontologisia tasoja, joista vain ylimmät ovat semanttista informaatiota, tietoa. Data-käsite sen sijaan viittaa raaka-aineeseen, joka voi olla hyvin erimuotoista. Usein data arkikielessä viittaa tietoteknisiin järjestelmiin tallennettuun informaatioon. Informaation ja datan välinen ero on hyvin suhteellinen.

Käytämme oppaassa pääasiassa termiä data (ja siitä muodostettuja yhdysanoja, kuten datalähde, datantuottaja, datakatalogi jne.) tarkoittamaan digitaalisessa muodossa olevaa raaka-ainetta. Käytännössä koko suomalainen alan keskustelu on kuitenkin puhunut tietovarannoista 1990-luvulta alkaen, joten käytämme myös termiä tietovarannot kun viittaamme yleiseen keskusteluun, sekä aiempiin raporteihin ja selvityksiin.

Datalla tarkoitetaan digitaalisesti tallennettua informaatiota, kuten dokumentteja, tietokantoja sopimuksista, transkripteja kuulemisista ja audio- sekä visuaalisia tallenteita tapahtumista. Vaikka ei-elektronisessa muodossa olevat resurssit, kuten fyysiset esineet (esimerkiksi vanhat paperidokumentit), eivät kuulu julkishallinnon avoimen datan piiriin, niin tällaisten resurssien saattaminen mahdollisuuksien mukaan digitaaliseen muotoon on kannustettavaa.

Oppaassa neuvotaan julkisen datan avaamiseen siinä nimenomaisessa mielessä, että data on raaka-ainetta, jota eri toimijat voivat muokata, yhdistellä, suodattaa ja jalostaa. Yhden toimijan tuottama lopputulos voi olla potentiaalisesti raaka-ainetta jollekin muulle toimijalle. Esimerkiksi Finlexin sisältämä ajantasainen lainsäädäntö ja suomi.fi-portaalin toimitettu sisältö ovat jalostettua informaatiota, mutta koneluettavassa muodossa ne voisivat olla raaka-ainetta sovelluskehittäjille.

Korostamme data-käsitteen avulla myös informaation tarjoamista digitaalisessa tietokoneiden avulla käsiteltävässä muodossa. Esimerkiksi Euroopan tietoverkkorikollisuutta koskevassa yleissopimuksessa data tarkoittaa tosiseikkojen, tietojen tai käsitteiden esitystä sellaisessa muodossa, että se soveltuu käytettäväksi tietojärjestelmässä, mukaan lukien ohjelmat, joiden avulla tietokone pystyy suorittamaan jonkin toiminnon (TEM 2009).

Datan ymmärtäminen raaka-aineeksi liittyy läheisesti seuraavaksi käsiteltävään ekosysteemi-käsitteeseen.

Data ekosysteemin osana

Hallinto tuottaa jatkuvasti suuren määrän dataa, joka voisi olla tehokkaamminkin käytössä huomioiden tietovarantojemme nykyinen laadukkuus. Asiaa lähestytään hahmottelemalla suomalaista avoimen datan ekosysteemiä. Aikaisemmissa keskusteluissa kansalaisia, yksityistä sektoria tai muita organisaatioita ei ole pidetty osapuolena tietovarantojen kehittämisessä, vaan ne on nähty ainoastaan tiedon loppukäyttäjinä. Hallinnon ohella myös kansalaiset, liike-elämä, järjestöt ja tutkimuslaitokset ovat merkittäviä toimijoita avointa dataa käsiteltäessä. Yksityisten toimijoiden tietovarannot eivät ole julkisia, mutta se ei estä näitä halutettavasta tuottamasta avointa dataa. Oppaassa esittämämme toimintatavat, työkalut, periaatteet ja suositukset ovat hyvin sovellettavissa kaikissa niissä organisaatioissa, kuten yrityksissä ja järjestöissä, jotka haluavat lisätä omien datavarantojen avointa käyttöä.

Oppaan yleistavoitteena on osaltaan edistää toimivan avoimen datan ekosysteemin syntymistä Suomeen. Siinä kansallisten tietovarojen hyödyntäminen lisääntyisi merkittävästi avoimuuteen pohjautuvan yhteistyön ja teknologian avulla. Ekosysteemillä tarkoitetaan yleisesti tietyn alueen toiminnallista kokonaisuutta. Voidaan puhua esimerkiksi tietyn järven tai metsän ekosysteemistä, joka koostuu alueen eliöiden ja elottomien ympäristötekijöiden dynaamisesta vuorovaikutuksesta. Avoimen datan yhteydessä ekosysteemi on monitasoinen ja -ulotteinen kokonaisuus, jossa raakadata on jakamisen ja jalostuksen osalta yhteistyön kohde. Raakadatalle ei käydä kauppaa, vaan se saatetaan helposti kaikkien toimijoiden saataville.

Datan hankkimisesta vapautuneet resurssit voidaan täysimittaisesti kohdistaa sen hyödyntämiseen. Laajamittainen datan hyödyntäminen tuottaa uusia palveluita, tutkimusta ja tietoa, josta osalla on kaupallista arvoa ja osa edistää demokratiaa, sivistystä ja ihmisten arkea ilman suoria taloudellisia kytköksiä. Lisääntynyt datan hyödyntäminen heijastuu positiivisesti datan tuottamiseen ja lisää jatkuvasti datavarantojen laatua ja käytettävyyttä. Ekosysteemissä hallinnon organisaatiot, kansalaiset ja yritykset ovat sekä datan käyttäjiä että myös sen tuottajia.

Internetin ja tietotyön yleistymisen myötä perinteisten liiketoimintamallien ja rahatalouden ulkopuolisen tuotannon merkitys on kasvanut, joita ilmentävät muiden muassa toimintamallit kuten open source, wikipedia ja sosiaalinen media. Oppaassa ajattelemme datan keräämistä, jalostusta, julkaisua ja uudelleenkäyttöä kokonaisuutena ja eri tekijöiden välisenä vuorovaikutuksena, eikä vain markkinoina ja vaihtokauppana. Ekosysteemiin liittyy mielikuva kokonaisuuden hyvinvoinnista ja toisaalta omien tarpeiden tyydyttämisestä osittain ekosysteemin rikkauden ja elinvoimaisuuden kautta.

Toinen käyttökelpoinen käsite aihepiiriin kokonaisvaltaiseksi ymmärtämiseksi on avoimen datan infrastruktuuri, joka niin ikään käsittää kaikki avoimen datan kanssa toimivat organisaatiot ja järjestelmät, toisin sanoen koko toimintaympäristön. Tämä malli sopii avoimen datan kentän jäsentämiseen esimerkiksi valtiollisella ja kunnallisella tasolla. Esimerkiksi laki paikkatietoinfrastruktuurista (ks. luku 3) kuvaa hyvin ajattelua, jossa ymmärretään tieto infrastruktuurin perusaineksena: "Paikkatietoinfrastruktuurilla tarkoitetaan saataville asetettuja metatietoja, paikkatietoaineistoja ja paikkatietopalveluja, verkkopalveluja ja -teknologioita, tietojen luovuttamista, saatavuutta ja käyttöä koskevia sopimuksia sekä koordinointi- ja seurantamekanismeja." Hallinnon rooli olisi tuottaa infrastruktuuria kaikkien käytettäväksi ja näin toimia mahdollistajana datan laaja-alaiselle hyödyntämiselle.

Ekosysteemillä korostamme teknisen järjestelmän ja institutionalisoitujen organisaatioiden ohella myös elävää, dynaamisesti muuttuvaa vuorovaikutusverkostoa, jonka osana hallinnon organisaatiot ja yksittäiset ihmisetkin ovat. Infrastruktuuri-ajattelu on oppaassa taustalla ja siihen tartutaan vahvemmin luvussa 6: *Avoimen datan infrastruktuuri*, jossa käsitellään organisaatorajat ylittäviä avoimen datan hankkeita, kuten kansallinen datakatalogi.

Tahtotila: Julkinen data avoimesti saataville ja hyödynnettäväksi

Datan avaamiseen voi suhtautua hyvin käytännönläheisesti. Oppaan taustaineiston sekä haastattelujen inspiroimina haluamme kuitenkin tuoda esiin laajemman näkökulman datan avaamiseen ja sen mahdollisuuksiin Suomen kannalta. Oheista visiota konkretisoivat oppaan lopuksi kootut toimenpide-ehdotukset. Vision tarkoituksena on edistää mielikuvaa siitä, mihin kaikkeen avoimen datan ajattelu voisi johtaa Suomessa. Visiomme on tarkoitettu kannustamaan lukijaa avoimen datan edistämiseen yleisellä tasolla, jotta aikanaan voitaisiin myös poliittisten linjausten avulla luoda toimenpideohjelmia tässä esitettyjen ja muiden tavoitteiden saavuttamiseksi.

Netinkäyttäjistä harvat miettivät, miten internetin hakukoneiden löytämä informaatio on tuotettu tai miten se on julkaistu verkkoon. Internetin kehitystrendeistä voi kuitenkin lukea, että yhä useammat ihmiset ja organisaatiot ovat alkaneet ajatella, että suuren osan omassa hallussa olevaa ja kertaalleen tuotettua dataa voisi tarjota maksuttomasti muille eikä se olisi keneltäkään pois.

Avoin data on meistä ensisijaisesti lähestymistapa ja asenne samalla tavalla kuin viime vuosina paljon käytetyt Web 2.0 ja sosiaalinen media -termit. Ne eivät ole yksiselitteisiä, kirjasta tai koulutuspäivästä omaksuttavia suoraviivaisia välineitä tai toimintaohjeita. Pikemminkin ne ovat kehityspolkuja, joita voi ottaa vähitellen käyttöön ja joiden avulla voi kehittää itselleen sopivia toimintarooleja tai soveltaa niitä organisaatioiden toimintaan (esim. Hintikka 2009; Kuntaliitto 2010). Samalla tavalla avoin data on yksi, mutta monia uusia mahdollisuuksia tuova lähestymistapa internet-aikakauteen.

Miten sitten pitää edetä ja toimia, jotta avoin datapolitiikka olisi meillä mahdollista? Olemme luoneet tulevaisuuskuvan tahtotilasta, joka on mahdollinen toteuttaa maassamme jo olevien resurssien, käytäntöjen ja toimintaympäristön ominaisuuksien avulla. Tarkastelemme seuraavaksi tätä tahtotilaa, joka sisältää viisi tavoitetta ja joita esittelemme esimerkkien avulla.

Tavoite 1: Uusia ratkaisuja kehittävät koulutetut suomalaiset

Yhteiskunnan positiivisen muutoksen tekijöitä ovat informoidut yksilöt. He tietävät, mitä yhteiskunnassa tapahtuu ja pystyvät ilmaisemaan itseään ja vaikuttamaan ympäristöönsä sekä yksilöinä että verkostojen ja järjestöjen osana. Julkishallinnon datan avoin jakelu lisää läpinäkyvyyttä ja erilaisten toimijoiden mahdollisuuksia osallistua yhteiskunnan kehitykseen.

Internetin yleistymisen myötä yksittäisten ihmisten mahdollisuudet saada tietoa yhteiskunnallisista ongelmista ja keskustella niiden ratkaisemisesta ovat parantuneet, esimerkiksi bloggaamisen ansiosta. Verkostoituneet yksilöt ovat osoittaneet kykenevänsä laajamittaisiin yhteistyöprojekteihin, joista on merkittävää päivittäistä hyötyä, kuten Wikipedia-tietosanakirjasta. Vähähiilisyttä edistävä *Porkkanamafia*-kuluttajaliike ja sähkömoottoreita bensa-autoihin asentava *Sähköautot - Nyt!* ovat onnistuneet tarttumaan ilmastomuutoksen uhkaan tavalla, joka muille toimijoille on ollut vaikeaa. Avoimen datan ympärille saattaa kehittyä samanlaisia ongelmanratkajayhteisöjä, sillä korkean koulutustason ansiosta meiltä löytyy paljon tietotyöhön tottuneita ihmisiä.

Kahden vuoden välein EU ministerit kokoontuvat sopimaan ministeritason sähköisen hallinnon julistuksesta (Ministerial Declaration on e-government). Se on keskeinen Eurooppalainen strateginen asiakirja. Yleensä tätä julistusta seuraa myös teollisuuden julistus. Vuoden 2009 marraskuussa EU ministereiden e-government kokouksessa Malmössä esiteltiin myös kansalaisten avoimesti sosiaalisen median välineillä koostama julistus (Eups20 2010). Tämä julistus syntyi aktiivisten kansalaisten aloitteesta, joukkovoimalla netin yhteisöllisillä ideointi-

(UserVoice) ja kirjoituslustoilla (MixedInk). Myös sen suomentaminen tapahtui vapaaehtoisvoimin Suomalaisissa sosiaalisen median yhteisöissä (alustoina Qai-ku, Twitter ja Etherpad). Olemme liittäneen julistuksen oppaan liitteeksi (liite 1) havainnollistamaan sitä kuinka aktiivisimmat kansalaiset kaipaavat yhteistyön tiivistämistä hallinnon kanssa.

Case: Asialista.fi

Asialista.fi on valmisteluvaiheessa oleva yhteisöllinen verkkoprojekti. Palvelu ko-koaisi kaikki yksittäiseen kunnallishallinnon päätöksentekoprosessiin liittyvät dokumentit ja hajallaan olevan tiedon yhdeksi kokonaisuudeksi wiki- ja blogi-ohjelmistoilla. Verkosto vapaaehtoisia aktiiveja voisi kirjoittaa yleiskielisiä tiivistelmiä kaupunginhallitusten asialistoista linkittäen niitä muihin verkossa oleviin aineistoihin. Tavoitteena on vähentää kunnallisen päätöksenteon seuraamiseen liittyviä *aikakustannuksia* (ks luku 4.5) ja vapauttaa kuntalaisten aikaa poliittiseen keskusteluun. Kun yksi ihminen kirjoittaa tiivistelmän yhteiseen palveluun, vapauttaa se aikaa muilta toisiin tehtäviin.

Palvelun kehittäjät ovat käyneet keskustelua kaupunkien tietotekniikkavastaavien kanssa koneluettavan rajapinnan avaamisesta asianhallintajärjestelmiin. Tällöin tietojen kerääminen Asialista.fi-palveluun voitaisiin automatisoida mahdollisimman pitkälle.

Tavoite 2: Talkoot ja yhteistoiminnan perinne

Avoin data auttaa laajentamaan informaatiopohjaisten hyvinvointipalveluiden tuotantoa hallinnon ulkopuolella. 2010-luvun hyvinvointiyhteiskunta on osin aktiivisten ja voimaantuneiden kansalaisten ja yhteisöjen itselleen ja toisilleen tuottama ja julkishallinto toimii mahdollistajana tarjoten maksuttomia dataresursseja.

Suomi on hyvinvointivaltio, jonka perustuslaissa on kirjattu valtiolle velvollisuus taata kansalaisille ihmisoikeuksien ja poliittisten oikeuksien lisäksi myös sosiaaliset oikeudet välttämättömään elintasoon ja huolenpitoon. Hyvinvointi ymmärretään Suomessa yksimielisesti käsittämään sekä henkisen, sosiaalisen, taloudellisen että ympäristön hyvinvoinnin. Ihmisten hyvinvoinnista huolehtiminen on synnyttänyt Suomeen laajan julkisen sektorin, joka suunnittelee ja suurelta osin toteuttaa yhteiskunnan infrastruktuuria ja palveluja. Hallinnon tuottaman datan osalta julkinen sektori on usein ainoa taho, joka jalostaa datasta arkea helpottavia ja hyvinvointia tukevia palveluita.

Suomessa on vahvoja yhteis- ja ilmaistoiminnan perinteitä, kuten talkookulttuuri, jokamiehen oikeus, kirjastolaitos sekä Linux- ja MySQL -ohjelmointihankkeet. Niitä voitaisiin hyödyntää nykyistä laajemmin myös kansalaisten ja julkishallinnon välisessä yhteistyössä ja julkisten palveluiden kehittämisessä (Mokka & Neuvonen 2006). Avoin data on yksi konkreettinen mahdollisuus saattaa yrityksiä, kansalaisia ja järjestöjä yhteen ja osaltaan tuottamaan yhteiskunnan hyvinvointipalveluita yhdessä julkisen sektorin kanssa.

Luvussa 1: *Elävä palvelu* käsittelemme hallinnon mahdollistajaroolia ja keinoja avoimeen yhteistyöhön

Case: Ymparisto.fi

Suomen luonnon seurantaan ja sen raportointiin käytetään nykyisin viranomaisytyötä noin 60 henkilötyövuotta. Vapaaehtoisten raportojien ja harrastajien joukkoutuneeksi (luku 3) työpanokseksi on arvioitu noin 200 henkilötyövuotta.

Luonnosta kiinnostuneiden ihmisten ja hallinnon edustajien vuorovaikutus toimii molempia osapuolia hyödyttäen ja havaintojen varmentaminen huomioiden. Ihmiset tuottavat raakadataa, jonka hallinnon työntekijät tarkistavat ja jatkavat arvioinnin ja korjaamisen jälkeen verkossa informaationa, johon mukana olevat vapaaehtoiset voivat reagoida ja kerätä sen pohjalta uutta raaka-ainetta tarjolle. Vastaavaa menetelmää käyttää muun muassa amerikkalainen tartuntatautien ehkäisykeskus Centers for Disease Control and Prevention (CDC). Tässä toimintatavassa kansalaiset lähettävät havaintojaan esimerkiksi sikainfluenssartunnoista ja CDC Emergency välittää niitä arvioinnin ja tarkistamisen jälkeen muun muassa Twitter-mikrokanavalla.

Tavoite 3: Laadukkaat tietovarantomme kansainvälisen profiloitumisen perustaksi

Suomalaiset tietovarannot ovat niin laadukkaita, että niitä esittelemällä ja hyödyntämällä on mahdollista profiloitua erityysosaajiksi maailmanlaajuisesti. Positiivinen maine antaa suomalaisille huippuosaajille ja organisaatioille mahdollisuuden toimia kansainvälisinä asiantuntijoina ja yhteistyötahoina kansainvälisissä hankkeissa.

Suomeen on aikojen kuluessa syntynyt laadullisesti tarkastellen maailmanluokan tietovarannot ja niiden potentiaalinen voima saadaan tehokäyttöön, kun niitä hyödynnetään mahdollisimman laajasti. Julkisen datan joustavaan tuotantoon, jakeluun, ylläpitoon ja käyttöön tarvitaan kansallisesti tekninen, juridinen ja organisatorinen infrastruktuuri. Tässä kehityksessä Suomi toivottavasti ottaa harppauksia eteenpäin.

Perusedellytys laadukkaiden tietovarantojen ylläpidolle on myös, että dataa keräävillä ja ylläpitävillä viranomaisilla on pysyviä resursseja. Vähintään yhtä tärkeää on, että ihmiset luottavat viranomaisiin ja siihen, että viranomaiset kunnioittavat tietosuojaa ja siihen liittyvää lainsäädäntöä. Luottamus auttaa kansallisten tietovarantojen ylläpidossa, koska ihmiset ja yritykset toimittavat rekisterinpitäjille kysytyt tiedot. Avoimuudesta ja maksuttomuudesta puhuttaessa monet tahot esittävät oikeutetusti huolensa korkean laadun säilymisestä. Resurssointi on taattava siinäkin tapauksessa, että dataa luovutetaan maksuttomasti. Avoin ja laaja jakelu ei saa johtaa datan tuottajan luotettavuuden heikkenemiseen esimerkiksi siten, että yksityisyydensuoja vaarantuu.

Luvussa 6: Avoimen datan infrastruktuuri käsittelemme keskeisiä teknisiä ja organisatorisia muutoksia, joilla Suomen tietoinfrastruktuuria voidaan kehittää..

Case: Väestölaskenta

Väestölaskenta on yksi esimerkki Suomen laadukkaista tietovarannoista. Väestölaskennan tiedot sisältävät asukasluvun lisäksi dataa esimerkiksi koulutuksesta ja työssäkäynnistä. Monissa maissa laskenta toteutetaan YK:n suositusten mukaan kymmenen vuoden välein lomakkeilla. Suomessa väestölaskenta tehdään vuosittain yhdessä yössä nappia painamalla ja ajamalla 37 rekisteriä yhteen. Tämän jälkeen väestölaskennan tiedot ovat hyvin nopeasti käytettävissä kaikille niitä tarvitseville organisaatioille hallinnon sisällä. Samaan pystyvät myös Tanska, Norja ja viimeisimpänä tulokkaana Ruotsi. Myös Hollannilla, Kanadalla, Australialla ja Uudella-Seelannilla on hyvä tietoinfrastruktuuri. Esimerkiksi Yhdysvalloissa, joka on julkishallinnon tietovarantojen avoimen jakamisen esimerkkinä, rekisterit ja tietovarannot eivät ole läheskään yhtä laadukkaita kuin Suomessa.

Tavoite 4: Dataa myös koneille

Datan avaamisesta käydyissä keskusteluissa koneluettavuuden ja avoimuuden panostaminen merkitystä on usein aliarvioitu. Datan koneluettavuuteen ja avoimuuteen panostamalla saataisiin jo nykyisin julkiset tietovarannot huomattavasti laajemmin ja monipuolisemmin käyttöön. Varsinaisen datan tarjoamisen rinnalla koneluettavuus on keskeinen elementti kokonaisvaltaisten hyötyjen saamiseksi.

Julkista dataa on jo nykyäänkin saatavilla hallinnon organisaatioiden verkkopalveluissa varsin runsaasti. Suuri osa aineistosta on kuitenkin sellaisessa muodossa, etteivät koneet pysty käsittelemään sitä. Datan ohjelmallinen ja automatisoitu hakeminen, käsittely ja yhdistely toisiin aineistoihin on työlästä tai käytännössä mahdotonta.

Esimerkiksi vireillä olevien kaavahankkeiden esittäminen kaupungin karttapalvelussa noudattaa julkisuuslakia. Useinkaan kaavojen paikkatieto ei ole

saatavilla koneluettavassa muodossa, jolloin lähiympäristön kaavahankkeiden automaattinen seuraaminen ei ole mahdollista.

Luvussa: 5 Tekninen valmistautuminen käsittelemme koneluettavuuden toteuttamista

Case: ReittiGPS

Älypuhelimissa toimiva ReittiGPS-sovellus on hyvä esimerkki koneluettavuuden ja automatisoinnin eduista verrattuna toistuvaan aikataulun selaamiseen netissä. Sovellus yhdistää kaksi datalähdettä: gps-anturin kertoman sijainnin ja Reittiopas-palvelun tietokannan tiedot joukkoliikenteen aikatauluista. Reittiopas-palvelun tietokannoissa on raakadataa julkisten kulkuneuvojen pysäkki-, reitti- ja aikataulutiedoista. Sovellus ottaa gps:sta paikkatiedon, asettaa sen lähtöpisteeksi ja kysyy käyttäjältä määränpäättä. Avoimen rajapinnan kautta ReittiGPS-sovellus hakee datan koneellisesti ja automatisoidusti kännykkään. Käyttäjälle kännykkäsovellus tarjoaa parilla napin painalluksella selkeän opastuksen, mille pysäkille hänen tulisi kävellä ja mihin kulkuneuvoon nousta..

Tavoite 5: Pienestä liikkeelle

Parhaiten saavutetaan tuloksia, kun kehityshankkeissa etsitään taustaksi tietoa muiden jo toteuttamista samankaltaisista hankkeista, opitaan tehdyistä virheistä ja lähdetään liikkeelle pienin askelin valmiina korjaamaan suuntaa heti kun siihen ilmenee tarvetta. On hyvä muistaa myös, että jonkun on oltava pioneeri ja avattava tietä.

Linux-käyttöjärjestelmän alkutahdit antavat hyvän esimerkin siitä, miten globaalit innovaatiot voivat nykyään käynnistyä. Linus Torvalds viestitti internetissä vuonna 1991, että hän on kehittänyt jotakin vielä keskeneräistä, joka saattaa kiinnostaa muitakin. Hän halusi kuulla muiden kehitysideoita, mutta ei luvannut toteuttaa niitä. Viestin varovaisesta sisällöstä huolimatta siitä muodostui alkusysäys erittäin merkittäväksi kasvaneelle yhteistyöprojektille. Lähes kaksikymmentä vuotta myöhemmin, avoimen lähdekoodin maksuton Linux-käyttöjärjestelmä on kehittynyt kilpailijaksi Windowsin ja Macin OS X:n rinnalle. Maksuttomuudesta huolimatta Linux on synnyttänyt merkittävää liiketoimintaa, jopa siinä määrin, että esimerkiksi IBM on luopunut oman käyttöjärjestelmänsä kehittämisestä, tarjoaa Linuxia ratkaisunaan ja edistää omaa liiketoimintaansa. Linuxin käyttöjärjestelmän ytimen arvoksi on laskettu miljardi euroa (García 2010).

Toivomme, että hallinnon organisaatiot samalla tavoin kertovat tietovaranoistaan, vaikka ne eivät vielä olisi valmiita avattaviksi ja ryhtyisivät vuoropuheiluun niiden käytön lisäämiseksi.

Luvuissa 1: *Elävä palvelu* ja 2: *Organisaation näkökulmia avoimuuteen* käsittelemme ketterää toimintamallia datan avaamisen kontekstissa.

Case: Linux

Lainaus Linus Torvaldsin lähettämästä sähköpostista, jossa hän esittelee ajatuksiaan uuden käyttöjärjestelmän ohjelmoinnista. Alkuperäisen viestin teknisiä yksityiskohtia on jätetty tässä pois.

Aug 26 1991, 8:12 am

Hello everybody out there using minix –

I'm doing a (free) operating system (just a hobby, won't be big and professional like gnu) for 386(486) AT clones. This has been brewing since april, and is starting to get ready. I'd like any feedback on things people like/dislike in minix, as my OS resembles it somewhat...

.... Any suggestions are welcome, but I won't promise I'll implement them :-)

Linus (torvalds@kruuna.helsinki.fi <mailto:torvalds@kruuna.helsinki.fi>)

**Monet toimivat nettipalvelut ovat
syntyneet kehittäjien omista tarpeista
ja heidän havainnoistaan.**



Datan avaaminen muiden käyttöön on paljon enemmän kuin teknologian soveltamista ja uusimista. Toimintaa suunniteltaessa on tarpeen huomioida alusta lähtien käyttäjät, toimintatavat ja oppimismahdollisuudet. Tässä luvussa käsittelemme avoimen datan teemaa ihmisten näkökulmasta. Tarkastelussa on, miten potentiaalisia avoimen datan käyttäjäryhmiä kohdataan, mitä on käyttäjälähtöinen innovointi (TEM 2010) ja mitä konkreettisia käyttökohteita avoimella datalla on. Tuomme keskusteluun myös *elävä palvelu* -käsitteen, jolla painotamme palvelun toiminnallisuutta.

Olemme jo johdantoluvussa esittäneet perusteluja datan avaamisen puolesta. Tavoitteena voi esimerkiksi olla rajallisten resurssien mahdollisimman tehokas käyttö, paremmin toimiva kansalaisyhteiskunta tai talouden kasvu. Datan avaaminen ei kuitenkaan automaattisesti johda näiden tavoitteiden toteutumiseen, vaan tavoitellut hyödyt saavutetaan vasta, kun avattua dataa myös käytetään.

Omaa dataansa jaettavaksi tarjoavalla organisaatiolla on monia keinoja, joilla se voi tukea todellisten *elävien palveluiden* syntymistä. Elävällä palvelulla tarkoitamme kaikkia niitä sovelluksia ja muita käyttökohteita julkishallinnon datalle, joiden avulla datasta saadaan irti hyötyjä. Datan avaaminen on siten keino eikä päämäärä itsessään, ja elävien palvelujen tuottaminen on sitä, mitä datan avauksella halutaan saada aikaiseksi. Useimmiten toivotaan, että syntyisi luovia mielenkiintoisia datan käyttötapoja, käytettäviä ja ennen kaikkea todellisiin tarpeisiin osuvia palveluita sekä niihin tukeutuvaa liiketoimintaa.

Uuden-Seelannin datanavausprojekteissa työskennellyt Nat Torkington kiteyttää artikkelissaan *«Lessons learned from the Open Data front lines»* (Torkington 2010) yhteisölähtöisen toimintatavan merkityksen uudenlaisten sovellusten syntymisessä. Helpoin tapa vakuuttaa ihmisiä avoimen datan hyödyistä on näyttää konkreettinen esimerkki, jonka ihmiset kokevat itselleen hyödylliseksi ja jossa avointa dataa aiotaan yhdistää uudella tavalla palvelun tuottamiseksi.

Case: Apps for Democracy (Washington DC.)

«Apps for Democracy kilpailu tuotti enemmän säästöjä Washington D.C:n hallinnolle, kuin mikään muu hanke.»

–Vivek Kundra, Washington D.C:n entinen tietohallintojohtaja, nykyinen Yhdysvaltain liittovaltion tietohallintojohtaja

Kenties maailman ensimmäinen kattava julkisten datavarantojen katalogi (ks. luku 6.1) on jo vuonna 2006 perustettu Washington D.C:n hallinnon ylläpitämä katalogi (<http://data.octo.dc.gov>). Katalogi sisältää satoja laadukkaita aineistoja reaaliaikaisista julkisen liikenteen syötetiedoista koulujen vertailutuloksiin ja seudullisiin tilastotietoihin. Katalogin julkistamisen jälkeen ei pariin vuoteen kuitenkaan ollut syntynyt moniakaan näkyviä sovelluksia, ja katalogi oli ollut lähinnä hallinnon itsensä käytössä. Apps for Democracy -kilpailu kehitettiin kannustimeksi, jotta katalogin aineistoja käytettäisiin laajasti.

Koko kilpailun organisointi maksoi kaupungille 50 000 dollaria, josta 20 000 jaettiin palkintorahoina. Tuloksena oli 47 toimivaa palvelua, joukossa niin mobiili- internet- kuin Facebook- ja Twitter-sovelluksia. Kaupungin tietohallinto-osaston laskelmien mukaan näiden sovellusten tuottaminen perinteisen hankintakäytännön kautta olisi maksanut yli kaksi miljoonaa dollaria. Suuri osa kustannuksista perinteisellä mallilla olisi syntynyt sisäisestä projektihallinnan ja hankintatoiminnan kustannuksista. Tietohallinto-osasto arvioi myös, että tavanomaisin menettelyin tällaisen sovellusmäärän saaminen kansalaisten käyttöön olisi kestänyt yli kaksi vuotta, kun siihen nyt meni vain pari kuukautta. Kilpailun järjestämisessä ja markkinoinnissa käytettiin maksuttomia sosiaalisen median työkaluja, joiden avulla kilpailun kohdeyleisö tavoitettiin tehokkaasti nopeassa ajassa.

(Ks. lisää 1.4.2 Innovaatiokilpailu kannustimena toimintaan)

1.1 Mihin avointa dataa voidaan käyttää?

Helposti mielletävien internetin kautta toimivien sovellusten ohella avoimen datan käyttökohteita on monia muitakin. Tässä jaottelemme käyttökohteet karkeasti neljään kategoriaan: **1) yhdistelmäpalvelut**, **2) koulutus, tutkimus ja tuotekehitys**, **3) prosessien automatisointi** ja **4) käyttäjien yhteistoiminnalliset tuotannot**. Lista ei ole kattava, mutta sen avulla voi alkaa miettiä oman organisaation luonnetta ja sen tuottaman datan eri mahdollisuuksia.

Yhdistelmäpalvelut

Oppaan kirjoittamisen aikaan arkea helpottavat verkkosovellukset olivat selkeästi yleisin tapa hyödyntää avointa dataa. Ne visualisoivat, suodattavat tai koaavat informaatiota eri lähteistä ja vastaavat loppukäyttäjien tiedontarpeisiin.

Mashup on yleisilmaisuuksia julkisille, liike-elämän tai kansalaisten perustamille avoimen datan sovelluksille, jotka koostavat eri lähteistä saatavilla olevaa dataa niillä periaatteilla, joita palvelun perustajat haluavat. Mashup-termille ei toisistaan ole vielä vakiintunutta suomennosta, mutta kutsumme sitä Tekniikan sanastokeskuksen suositusten mukaisesti yhdistelmäpalveluiksi.

Vuonna 2005 amerikkalaistoimittaja Adrian Holovaty kyllästyi soittamaan joka aamu poliisilaitokselle kysyäksään rikosraporttia. Kun hän havaitsi, että raportit ovat saatavilla automaattisesti, hän yhdisti ne tuolloin vasta julkaistuun Google Maps -sovellukseen. Tuloksena oli ChicagoCrimes-palvelu, josta näki lähes reaaliaikaisesti, millaisia rikoksia on tapahtunut. Suomessa vastaavan sovelluksen teki Tilannehuone.fi, joka yhdisti Hälytyskeskuksen tilannetiedotuksen samaiseen Google Maps:iin. Yksinkertainen ja kiinnostava Tilannehuone.fi kiinnostaa, sillä siinä vierailee viikoittain lähes 300 000 kävijää. Yhdistelmäpalvelut sinänsä voivat vastata monenlaisiin tarpeisiin, kuten arjen helpottamiseen, politiikan läpinäkyvyyden lisäämiseen tai, kuten Adrian Holovaty tapauksessa, olla journalistisen työn väline.

Koulutus, tutkimus ja tuotekehitys

Tutkimusorganisaatioissa datan helppo saatavuus tukee laadukkaan tutkimuksen syntymistä. Koulutuksessa dataa voidaan hyödyntää asioiden havainnollistamiseen. Loistava esimerkki visualisaation voimasta on Tukholman Karoliinisen Instituutin professorin Hans Roslingin johdolla kehitetty GapMinder-palvelu. Ohjelmiston esiaste syntyi tarpeesta visualisoida opiskelijoille, kuinka 1960-luvulla syntynyt käsitys maailman kahtiajakautuneisuudesta perheeseen ja eliniän odotteen perusteella ei enää pidä paikkaansa.

Kuluttajille suunnattuja yhdistelmäpalveluita suuremmassa mittakaavassa tiedon louhintaa, yhdistelyä ja visualisointia voidaan tehdä tutkimus- ja tuotekehitysprojekteissa. Tällöin tavoitteena ei ole niinkään yksittäinen arkea helpottava tai hallinnon läpinäkyvyyttä lisäävä yhdistelmäpalvelu, vaan kokonaan uuden tiedon synnyttäminen tai jonkin toiminnan optimointi laajaan data-aineistoon perustuen. Esimerkiksi liikennemittausten, julkisen liikenteen käyttötilastojen ja erilaisten alueita koskevien tilastojen perusteella voidaan tehdä kaupungin liikennejärjestelmää koskevia optimointimalleja tai vaikkapa yksityisen toimijan toimesta optimoida yrityksen palvelupisteverkostoa. Nykyisin erilaiset organisaatiot tekevät tämän tyyppisiä optimointeja ja simulaatioita nojautuen omiin datavarantoihinsa, mutta avoin data mahdollistaisi myös lisälähteiden ja muiden organisaatioiden datavarantojen hyödyntämisen.

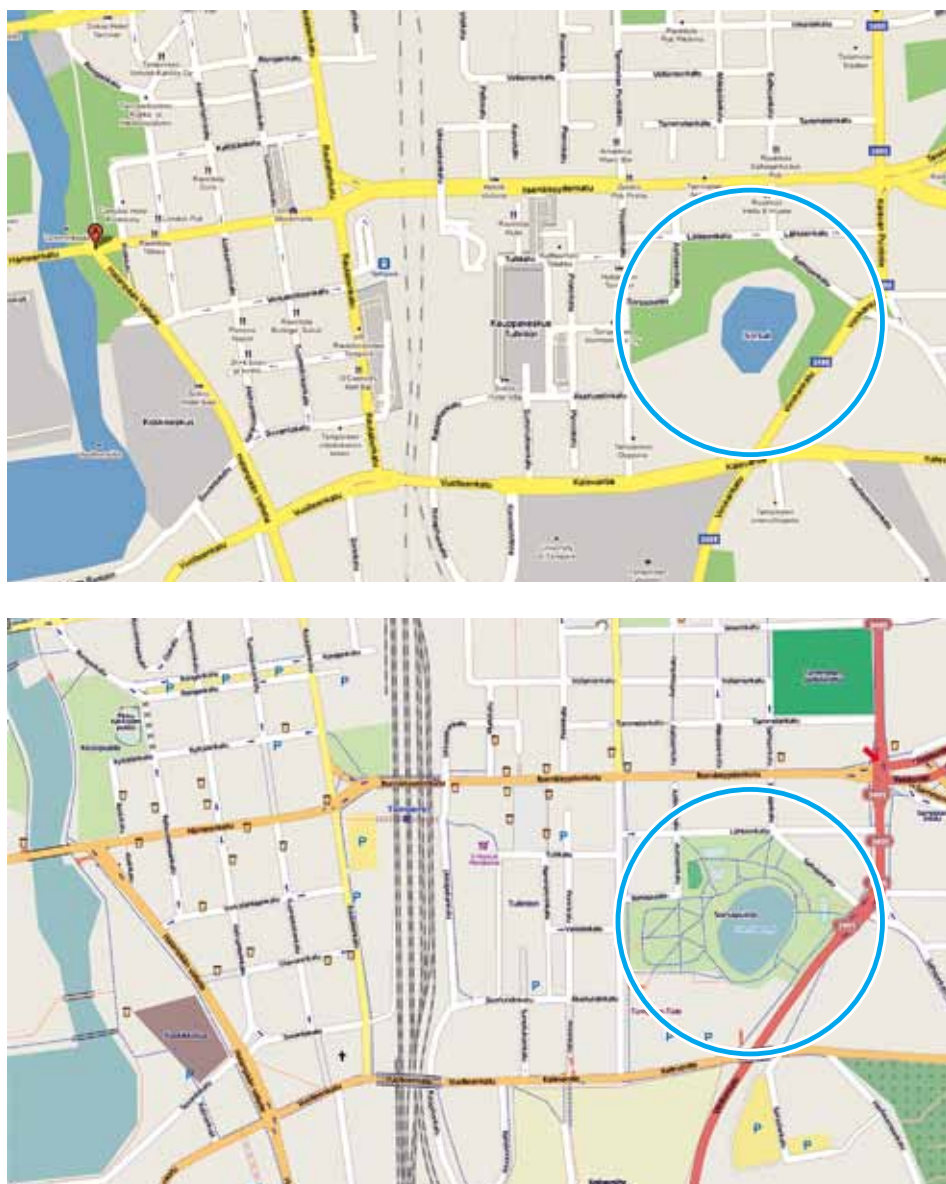
Automaatio

Dataa voidaan hyödyntää myös automaatiossa, jossa datan avulla ohjataan tai helpotetaan jotakin prosessia. Esimerkiksi erilaisten nettipalveluiden yhteydessä osoitelmakkeiden täydentämistä ja tarkistamista voidaan automatisoida postinumero- ja osoitedatan avulla. Voidaan ajatella myös lämmitys- ja ilmastointijärjestelmiä, jotka hyödyntäisivät säädataa, sekä sähköverkon kapasiteetista kertovaa dataa ja automaattisesti ohjaisivat itseään siten, että sähkön kulutus laskisi ja etenkin kulutushuiput tasaantuisivat (Smart Grid). Tämä on vielä laajasti toteutumatta, mutta tämänkin suuntainen kehitys voisi nopeutua, mikäli lisädataa olisi saatavilla.

Käyttäjien yhteistoiminnalliset tuotannot

Oma lukunsa avoimen datan hyödyissä on aineiston laadun parantaminen ja kerääminen kollektiivisesti sekä päällekkäisen työn vähentäminen yhteiskäyttöisen datavarannon myötä. Jeff Howe (2006) kehitti termin *crowdsourcing* kuvaamaan internetin mahdollistamia uusia tapoja organisoida työtä. Termi tarkoittaa pelkistettynä perinteistä työn ulkoistamista (*outsourcing*) internetiin nimeämättömälle ihmisjoukolla (*crowd*). Termille ei ole vielä vakiintunutta suomennosta, on puhuttu esimerkiksi talkoistamisesta ja joukkouttamisesta. Crowdsourcing'issa on monia eri muotoja. Tyypillisesti joko etsitään parasta mahdollista ratkaisua, jota voi ehdottaa periaatteessa kuka tahansa, tai dataa kerätään, luokitellaan, lajitellaan, tuotetaan ja jalostetaan kollektiivisesti.

Wikipedia ja OpenStreetmap ovat kenties tunnetuimpia esimerkkejä crowdsourcing-mallin toiminnasta, mutta jostain syystä usein unohdetaan, että kansalaiset ja yritykset voivat tuottaa dataa myös julkishallinnon hyödynnettäväksi. Kotimainen esimerkki hallinnon hyödyntämisestä crowdsourcing'ista on Helsingin kaupungin pilottipalvelu Fillarikanava, jossa pyöräilijät merkitsevät kartalle arkisia pyöräilyoloihin liittyviä havaintojaan, jotka tallentuvat paikkatiedoksi, joka voidaan huomioida alueita koskevilla suunnittelu- ja parannushankkeissa. Hieman erilainen sovellus on Britanniassa toteutettu opetusaineistojen yhteistuotanto- ja jakelualusta *Times Educational Supplement (TES)*. Sen on arvioitu säästävän opettajien työaikaa miljardin punnan edestä kahdessa vuodessa (UK 2009), kun opettajat ovat alkaneet palvelun kautta jakaa aineistoja keskenään sen sijaan, että kaikki tuottaisivat oman aineistonsa saman opetussuunnitelman mukaisten sisältöjen opettamiseen.



Kuva 1.1: Yhteisöllisesti ja vapaaehtoisvoimin tuotettu Open Street Map on mm. Tampereen keskustan jalankulkuväylien ja puistojen, sekä nimistön osalta huomattavasti Google-karttaa tarkempi. Suomen osalta yhteisö on piirtänyt ja GPS-paikantimilla mitannut vapaaehtoisesti kartan alusta asti. Open Street Map -yhteisölle voidaan tehdä myös paikkatietolahjoituksia, kuten on tapahtunut Yhdysvalloissa.

1.2 Käyttäjälähtöinen innovointi

Palveluiden elinkaari alkaa yleensä tarpeiden tunnistamisesta, mitä seuraa ideointi, testaus ja käyttökokemusten ja palautteen perusteella tapahtuva iteratiivinen kehitystyö. Käyttäjälähtöinen tuotekehitys ja innovointi ovat malleja, joissa palvelun tuottaja ottaa loppukäyttäjät mahdollisimman varhaisessa vaiheessa mukaan ohjaamaan palvelun ideointia ja kehitystä.

Edellisessä jaksossa käsitelimme erilaisia tapoja hyödyntää avointa dataa. Yhtä oikeaa tapaa ei ole ja samaa dataa voidaan käyttää monin eri tavoin. Lopulta toimivimmat tavat voivat nousta miltä tahansa aineistoa käyttävältä taholta, kuten kansalaisjärjestöistä, yrityksistä, yksittäiltä ohjelmoijilta ja idearikkailta kansalaisilta, toisista virastoista ja laitoksista jne. Suurella todennäköisyydellä parhaat tavat hyödyntää jonkin organisaation tuottamaa dataa syntyvät aivan jossain muualla kuin kyseisen organisaation sisällä.

Se, kuinka toimiva tai hyvä jokin sovellus on, riippuu sovelluksen käyttäjien tarpeista. Monet toimivat nettipalvelut ovat syntyneet kehittäjien omista tarpeista tai heidän havainnoistaan. Tällöin käyttäjät ovat löytäneet tarjoumasta (affordanssi) itselleen sopivan palvelun. Päinvastaisiakin esimerkkejä löytyy: on perustettu palveluja, jotka ovat tyylikkäitä ja toimivia, mutta jotka eivät kiinnosta käyttäjiä. Usein tällaiset epäonnistumiset olisi onnistuttu välttämään tutkimalla etukäteen käyttäjien tarpeita ja heidän osaamistaan palvelujen käytössä.

Avoin ekosysteemi mahdollistaa sellaisen toimintatavan, että ihmiset, jotka tarvitsevat jonkin palvelun, voivat itse alkaa ideoida, luonnostella ja toteuttaa sitä. Tämä on hyvä malli kehittää palveluita, jotka vastaavat todellisia tarpeita. Massojen ohella myös pienet ryhmät ja jopa yksittäiset ihmiset voivat saada aikaan omiin tarpeisiinsa vastaavia palveluita. Toisaalta avoimen datan tarjoajan näkökulmasta on edullista, mikäli suuri joukko käyttää tarjottuja datavarantoja. Esimerkiksi verkkokirjakauppa Amazon tarjoaa valtaisan osto- ja liiketoimitietokantansa maksutta ohjelmoijien käyttöön. Ohjelmoijat voivat hyödyntää aitoja tietokantoja ja saavat referenssimahdollisuuksia työuralleen tai yhtiölleen, kun taas Amazon saa maksutta kymmeniä tuhansia ohjelmoijia tekemään uusia innovatiivisia käyttöliittymiä itselleen. Sovellukset lisäävät Amazonin verkkonäkyvyyttä, kauppatapahtumia ja myyntiä.

Mikäli data ei ole avointa, vastuu ideoinnista ja toteutuksesta on dataa tuottaneella organisaatiolla. Julkishallinnon toimijoiden tehtäviin kuuluu tuottaa datansa perusteelta yleisimpiin tarpeisiin vastaavat niin sanotut perusnäköymät. Erikoisemmat ja pienempien ryhmien tarpeet jäävät väkisin toteuttamatta. Esimerkki edellä mainittu ChicagoCrimes-yhdistelmäpalvelu kuvastaa hyvin tällaista toimintaa. Palvelulle oli olemassa selkeä tarve, mutta se ei todennäköisesti olisi tullut poliisilaitoksen tietopalveluyksikön ideapalavereissa esille. Vaikka idea olisi syntynyt, niin rikostoitimittajat olisi tulkittu niin pieneksi kohderyhmäksi, ettei poliisilaitoksen resursseja todennäköisesti olisi kohdistettu palvelun tuotantoon.

Käyttäjälähtöistä tuotekehitystä äärimmillään on, kun palvelun kehittäjät ovat itse myös loppukäyttäjiä, esimerkiksi YouTuben sanotaan syntyneen sen perustajien tarpeesta jakaa netissä videotiedostoja (Cloud 2006). Samoin ensimmäinen wiki-sovellus syntyi, kun Ward Cunningham vuonna 1995 loi WikiWiki-Web-alustan, jotta ohjelmoijat voisivat helposti jakaa ohjelmakoodiin liittyvää tietoa keskenään. Julkishallinnon avoin data mahdollistaa sen, että ihmiset voivat itse omia tarpeitaan varten laajentaa julkisia palveluita. Esimerkiksi julkisilla liikennevälineillä keskustaan suuntautuvien työmatkojen matka-aikoja ja asuntojen hintoja kartalla visualisoiva *Mapumental*-palvelu syntyi asunnon ostoa suunnittelevien ihmisten tarpeista. Vastaavasti Hilmappi-palvelun tuottanut suomalainen Gemilo haistoi markkinaraon yritysten tarpeessa seurata helposti julkisia tarjouskilpailuja alueellisesti.

1.3 Ketterät menetelmät datan avaamisessa

Nyt digitaalisuuden ja globalisaation myötä organisaatioiden toimintaympäristö ja sitä kuvaava informaatiota muuttuu nopeammin kuin kukaan tai mikään organisaatio pystyy käsittelemään. Murros kaipaa orientoitumiseen uudenlaisia lähestymistapaa – ketteryyttä. Toimintaympäristön jatkuvasti muuttuessa ei välttämättä ole järkevää suunnitella yksityiskohtia kovin kauas tulevaisuuteen. Mitä kauemmas tulevaisuuteen suunnitellaan, sitä hatarampiin oletuksiin suunnitelma perustuu.

Ketteristä toimintatavoista on pitkään puhuttu ohjelmistokehityksen piirissä, mutta niiden logiikkaa voidaan soveltaa muunkinlaisiin prosesseihin. Ketteryys ei tarkoita kaiken nopeuttamista, kiihdyttämistä ja kiirettä, vaan jatkuvaa ohjautumista. Oleellista on, että päätöksiä tehdään jatkuvasti päivittyvän tiedon varassa – muuttuviin olosuhteisiin reagoiden ja tekemisen aikana opittua hyödyntäen. Asioita ei pyritä saattamaan kerralla valmiiksi, vaan niissä voidaan edetä jatkuvasti pienin askelin ja uusien iteraatiokierrosten kautta.

Ketteryydestä tai iteratiivisuudesta puhuttaessa korostetaan tarpeita vastaavien ja toimivien ratkaisujen ensisijaisuutta sekä kehittäjien välistä tehokasta viestintää. Toiminnassa voidaan olla mukana proaktiivisesti ja sitä kautta määrittellä toimintaympäristöä. Tämänkaltaisen toimintakulttuuri ei lamaannu siitä, ettei tulevaisuutta voida ennakoita kovinkaan pitkälle. Iteratiivisuus tarkoittaa myös virheiden nopeampaa korjausta. Kun toiminta etenee pienemmissä paloissa ja sitä arvioidaan koko ajan, myös mahdollisten virheiden korjaaminen ja suunnanmuutokset ovat helpompia. Viime kädessä ketteryyttä voidaan ajatella vuorovaikutuksena. Ketterä toimija operoi vuorovaikutteisesti ympäristönsä – ekosysteeminsä – kanssa.

Vielä on liian varhaista arvioida Suomen tai maailman mittakaavassa, mihin julkishallinnon datavarantojen avaaminen johtaa. Myöskin tämä opas esimerkkeineen ja toimintaohjeineen on tuomittu vanhenemaan nopeasti. Ketterät menetelmät soveltuvatkin siksi hyvin lähtökohdaksi julkishallinnon organisaatioiden datan avaamiseen. Ketteryyteen viitataan seuraavissa luvuissa, joissa esitellään datavarantojen inventaariota ja julkaisun teknologioita ratkaisuja. Ideana on, että liikkeelle kannattaa lähteä pienestä. Ensimmäiset datan avaamisen myöstä syntyneet elävät palvelut viitoittavat tietä jatkoon.

1.4 Organisaatio mahdollistajana

Toimintaympäristön muuttuessa myös organisaatioiden roolit muuttuvat. Tämä näkyy jo nyt esimerkiksi sisältöteollisuudessa internetin kehityksen myötä. Äänitemyyntin mullisti tietokonevalmistaja Apple iTunes-verkkokaupallaan eikä musiikkiteollisuus. Google on syönyt osan kaupallisten tietokoneohjelmistojen markkinoista maksuttomilla palveluillaan. Positiivisesti katsoen muutos mahdollistaa esimerkiksi hallinnolle uudentyyppistä toimintaa, joka olisi ollut ennen internetiä liian kallista. Uusia rooleja ovat esimerkiksi seuraavat:

- Informaationtarjoaja
- Mahdollistaja (enabler)
- Alustan tarjoaja (platform)
- Fasilitoija / koordinoija
- Konsultoiva läsnäolo verkossa (Hintikka 2009).

Johdannossa kuvattiin avoimen datan ekosysteemiä, jossa on lukuisia toimijoita eri rooleissa ja data on yhteistä uusiutuvaa raaka-ainetta. Avoimen datan ekosysteemissä hallinto toimii ainakin informaation tarjoajana ja mahdollistajana. Kun dataa tarjotaan maksuttomasti vapaaseen käyttöön, mahdollistuu myös roolien jalostuminen ja erikoistuminen. Asetelman voi nähdä niinkin, että avoin data on informaatioyhteiskunnan infrastruktuurin rakentamista. Se mahdollistaa korkeamman jalostusasteen palvelujen kehittämisen sosiaalisia innovaatioita myöten.

Vuoden 2008 aikana Yleisradion hallitus täydensi yleistä yhtiöstrategiaa mahdollistajastrategialla (2008). Mahdollistajastrategian keskeinen viesti on, ettei organisaatio enää vain toteuta omaa toimintaansa vaan osaltaan tarjoaa myös muille mahdollisuuksia aktiivisuuteen.

Tim O'Reilly (2010) kuvaa hallinnon muuttuvaa paradigmaa verkkoteknologioiden ja avoimen datan maailmassa termillä «government as a platform» – hallinto alustana. Ideana on, että hallinnon tulisi keskittyä infrastruktuurin tekemiseen ja siten mahdollistaa kestävänsä yksityisten toimijoiden ekosysteemin kehittymisen. Toimintamallina tämä ei ole hallinnossa mitenkään uusi, vaan esimerkiksi tieverkoston tekeminen on mielletty luontaiseksi hallinnon organisoi-maksi toimenpiteeksi (sen sijaan, että hallinto organisoisi ihmisten ja tavaroiden kuljetuksia), jonka myötä monenlainen yksityinen teihin ja liikkumiseen liittyvä toimeliaisuus on mahdollistunut. Avoimen datan kohdalla tämä tarkoittaa, että pienentämällä hallinnon roolia informaation välittämisessä informaatio saattaisi jopa välittyä paremmin. Nykyisin monet hallinnon organisaatiot pitävät informaation välittämistä omien verkkosivujen kautta tärkeämpänä kuin teknistä avoimen datan infrastruktuuria, joka mahdollistaisi, että joku muu voisi käyttää dataa.

Tim O'Reillyn ajatuskulkua Suomen oloihin sovittaen puhumme hallinnosta mahdollistajana, joka on hieman enemmän kuin vain pelkkä alusta. Mahdollistajuus on aktiivista toimintaa, kun alusta antaa mielikuvan passiivisesta. Avoimen datan suhteen sanomme, että hallinnolle olisi perusteltua tarjota dataa uudelleenkäytettävässä muodossa ja toimia alustana, mutta sen lisäksi aktiivisesti kannustaa yrityksiä, kansalaisia ja muita tahoja hyödyntämään dataa.

1.4.1 Avoin toiminta ja viestintä datan hyödyntäjille

Datan avaamisen projektit eivät ole irrallisia teknisiä hankkeita, vaan ne kannattaa rakentaa dataa käyttävien ja jalostavien ihmisten ja yhteisöjen ympärille osaksi avoimen datan ekosysteemiä. Aiemmin esiteltyjä käyttäjälähtöisen innovoinnin periaatteita voidaan soveltaa tiedonavaamisprosessiin ja datarajapinnan luomiseen. Datan käyttäjät ovat parhaita asiantuntijoita siinä, kuinka dataa kannattaa tarjota, jotta sitä aktiivisesti hyödynnettäisiin ja eläviä palveluita syntyisi. Dataa avaavat organisaatiot voivat esitellä rajapintoja, saada palautetta ja saada ihmiset auttamaan niiden kehityksessä.

Datan julkaisussa olennaista on pyrkiä tekemään datasta mahdollisimman löydettävää ja tavoittaa tavalla tai toisella juuri kyseisestä datasta kiinnostuneet tahot. Löydettävyyttä voidaan parantaa teknisesti muun muassa linkitetyn datan ajatustavan mukaisesti, hakukoneoptimoinnilla ja viemällä data suurempiin ohjelmoijien tuntemiin datakatalogeihin. Olennainen tekijä datan löydettävyydessä on kuitenkin kertoa sen olemassaolosta myös perinteisemmän promomisen mielessä. Ensimmäinen näkyvyys saanut sovellus tuo näkyvyyttä itse datalähteelle ja uusia sovelluksia syntyy sen jälkeen nopeammin. Avattujen tietovarantojen hyödyntämistä nopeuttaa, mitä aikaisemmin ja laajemmin kehittä-

miprosesseihin saadaan mukaan datavarantoja käyttäviä tahoja. Haasteena on laskea osallistumisen työläyttä ja organisoida yhteistyö tehokkaasti – niin että se tuottaa enemmän kuin vie osallistujiltaan.

Viestinnän kannalta keskeisessä roolissa ovat ohjelmoijat, jotka käyttävät dataa ja jalostavat niistä palveluita. Ohjelmoijat voivat työskennellä yrityksissä tai tutkimuslaitoksissa tai ohjelmoida vapaa-aikanaan. Tietenkään ohjelmointitaito ei ole vaatimuskriteeri avoimen datan yhteisössä mukana olemiselle. Kaikkien näkemyksiä datan käyttökohteista konkreettisten ongelmien ratkaisuun tarvitaan riippumatta siitä, osaako näkemyksen omaava henkilö ohjelmoida vai ei. Ohjelmoijien rooli on kuitenkin siksi merkittävä, että datan muuttaminen informaatioksi käytännössä aina vaatii käyttöliittymien tekemistä, ohjelmallista käsittelyä, visualisointia, yhdistelyä tai muunlaista havainnollistamista.

Datan avaamista voidaan valmistella tavanomaisista suljetuista projekti-käytännöistä poiketen osittain avoimesti internetissä. Luonteva tapa tavoittaa ohjelmointiyhteisöjä ja muita datan hyödyntämisestä kiinnostuneita tahoja on jalkautua sosiaalisen median palveluihin. Muun muassa Sitra on koordinoanut jo muutaman vuoden avoimessa internetissä erilaisia kansalaispaneeleja niin strategiatyössä kuin uusien strategia-aiheiden etsinnässä. Aiemmin suljetut, asian-tuntijakeskeiset tapahtumat on avattu internetiin kenen tahansa kiinnostuneen ulottuville (Luoto 2010). Monet muutkin organisaatiot ovat jo löytäneet sosiaalisen median. Eri tutkimusten mukaan hallinto on ollut Suomessa jopa liike-elämää aktiivisempi internetin uusien mahdollisuuksien käyttöönotossa. Läsnäolo internetissä mahdollistaa tyystin uuden kontaktipinnan, jota ei aiemmin ole ollut. Tästä esimerkkeinä vaikkapa poliisin toiminta IRC-Galleriassa, Facebookissa sekä netari.fi nuorisotoiminta useissa eri verkkopalveluissa.

1.4.2 Innovaatiokilpailu kannustimena toimintaan

Teknisen mahdollisuuden lisäksi elävien palveluiden ja sovellusten syntymiseen tarvitaan myös yhteisöjä: toimijoita, ohjelmoijia, tutkijoita, yrityksiä, järjestöjä, jne, jotka tarttuvat tilaisuuteen ja käyttävät dataa luovasti raaka-aineena. Mitä kiinteämpi ja keskusteleavampi datan kanssa toimivien ihmisten ja organisaatioiden ekosysteemi Suomeen kehittyy, sitä suurempi hyöty aukenevista datavarannoista saadaan Suomeen ja suomalaisille. On monia tapoja, joilla organisaatiot voivat varsin kustannustehokkaasti tukea avoimen datan hyödyntämistä ja samalla kerätä kokemusta yhteisöjen kanssa toimimisesta. Esimerkiksi oppaan kirjoittajilla oli mahdollisuus osallistua Helsingin seudun liikennelaitoksen järjestämään Reittioppaan API:n käyttäjien ja kehittäjien yhteiseen työpajaan. Tässä esitetään hieman laajemmin datan hyödyntämiseen tähtäävät innovaatiokilpailut, jotka ovat etenkin ekosysteemin muotoutumisen alkuvaiheessa toimivaksi havaittu konsepti.

Luvun alussa kerroimme Washington D.C:n datakatalogista ja sen hyödyntämiseksi järjestetystä Apps for Democracy -kilpailusta. Vaikka Vivek Kundra lausunto ja Washington DC:n tietohallinnon taloudelliset laskelmat kilpailun hyödyistä ovat inspiroivia, niin suoraan ei voida sanoa kilpailun tuoneen kahden miljoonan dollarin säästöjä kaupungille. Suurinta osaa syntyneistä palveluista kaupunki ei olisi normaalioloissa hankkinut eivätkä kilpailun kautta syntyneet palvelut muutoinkaan ole suoraan vertailukelpoisia kaupungin hankintojen kanssa, vaikka ne ovatkin kaupunkilaisille hyödyllisiä. Siitä ei kuitenkaan ole epäilystäkään, että kaupunki mahdollisti näiden lukuisten hyödyllisten verkkopalveluiden syntymisen kaupunkilaisille, sai runsaasti oppia yhteisöjen kanssa toimimisesta sekä positiivista julkisuutta varsin kohtuullisin kustannuksin.

Monet kilpailuun osallistuneista sovelluksista ovat jääneet elämään ja niitä

on kehitetty ja joidenkin ympärille on perustettu liiketoimintaakin. Kauaskantoisin ja merkittävin hyöty oli, että kilpailu loi pohjaa Washingtonin avoimen datan ekosysteemille. Nykyisin kehittäjät tuntevat Washingtonin datakatalogin ja muut avoimen datan toimijat samaan tapaan kuin he tietävät, että Googlella on karttarajapinta, jota voi vapaasti hyödyntää. Kilpailu oli alkusysäys Washingtonin avoimen datan ekosysteemille, joka koostuu datan lisäksi aktiivisista toimijoista.

Innovaatiokilpailun ideana oli antaa data kansalaisten ja yritysten käsiin ja palkita parhaiden sovellusten kehittäjiä rahallisesti ja julkisuudella. Vastaavia avoimen datan hyödyntämiseen tähtääviä innovaatiokilpailuita on järjestetty ympäri maailman ja vuonna 2009 myös Suomessa nimellä *Kansalaisosallistujan työkalut – Apps for Democracy Finland*.

Kansalaisosallistujan työkalut -kilpailu sai jo järjestelyvaiheessa erittäin positiivisen vastaanoton. Kilpailuun tuli 23 ehdotusta, joista noin puolet oli toimivia sovelluksia ja puolet konseptitason ehdotuksia (ks. Liite 2). Muutamia kilpailuun osallistuneita ideoita on sittemmin jatkokehitetty (ainakin Veropuu, Asialista, Nomen est Omen ja Simupedia, Eduskunnan kyselyportaali) ja muutkin ehdotukset ovat herättäneet kiinnostusta. Aivan kuten Washingtonin kilpailussakin kenties pysyvin vaikutus on orastavassa suomalaisessa avoimen datan ekosysteemin syntymisessä, jonka kehittymiseen kilpailuun osallistuneet hallinnon organisaatiot ja kilpailijat ovat sittemmin vaikuttaneet. Vuonna 2010 suomalainen Apps for Democracy Finland -kilpailu järjestetään Verkkodemokratiaseuran vetämänä.

**Tavoitteena on saada kokemuksia
datan avaamisesta normaalia
tietojärjestelmien uusimissykliä
nopeammin.**

2. Organisaation näkökulmia avoimuuteen

Datan avaamiseen ollaan Suomessa tarttumassa nyt laajalla rintamalla. Liikenne- ja viestintäministeriön on asettanut työryhmän edistämään julkisen informaation saatavuutta - valtiovarainministeriö kaksi (LVM & VM 2010). Valtakunnallinenkin media on tarttunut asiaan (HS 2.7.2010). Mitä tämä tarkoittaa yksittäisen organisaation kannalta?

Päätökset datavarantojen avoimesta ja maksuttomasta jakamisesta liittyvät taloudellisiin seikkoihin, vallankäyttöön, juridiikkaan ja tekniikkaan. Hyvin usein mukana on laaja toimijaverkosto, jolloin yhteisten päätösten tekeminen on aikaa vievää. Kansainvälisiä esimerkkejä - niin valtioiden, kaupunkien, kuin yksittäisten organisaatioidenkin - avoimeen dataan liittyvistä strategisista päätöksistä on runsaasti.

Oppaan kirjoittamisen aikaan yleinen ilmapiiri hallinnon datan avaamista kohtaan on nopeasti muuttunut positiivisemmaksi, mutta aihetta ei ole suoraan mainittu vielä monienkaan suomalaisten organisaatioiden strategioissa tai muissa julkilausutuissa tavoitteissa. Yleisemmällä tasolla esimerkiksi kansallisessa tietoyhteiskuntastregiassa (VNK 2006) tai EU-lainsäädännön kautta (Inspire ja PSI) aihe on kuitenkin jo heijastunut useimpien julkishallinnon organisaatioiden toimintaan.

Avoimen datan julkaiseminen tukee monien julkishallinnon organisaatioiden strategisia tavoitteita, kuten edistää kansalaisten osallistumismahdollisuuksia ja yritysten liiketoimintamahdollisuuksia. Useimmiten datan avoin ja maksuton tarjoaminen linkittyy suoraviivaisimmin viestintä- ja tietohallintostrategioihin. Tämä kuitenkin riippuu hyvin paljon organisaatiosta. Esimerkiksi Tilastokeskuksen tehtävä on tiedon tuottaminen, jolloin avoimen datan tarjoaminen liittyy suoraan sen strategiaan. Sen sijaan esimerkiksi TEKESin ja Finpron kaltaisten organisaatioiden tavoitteet liittyvät suomalaisen yritystoiminnan edistämiseen. Tällöin suhde dataan on välillisempi.

Yleensä pyrkimyksenä on, että organisaation muutoinkin tuottamille tietovarannoille saadaan avoimuuden myötä lisää käyttöä, joka tavalla tai toisella hyödyttää myös organisaatiota. Avoimen datan ekosysteemi ei kuitenkaan toimi tarkan vastavuoroisuuden perusteella, vaan hyödyt tulevat osittain epäsuorasti ekosysteemin kokonaisuuden kautta. Esimerkkejä epäsuorista hyödyistä ovat organisaation tunnettavuuden kasvu ja sen tarjoamien palveluiden kysynnän kasvaminen.

Yksittäiset julkishallinnon organisaatiot ovat datan avaamisessa aloitteentekijän asemassa. Useimmissa organisaatioissa resurssit datan avaamiselle eivät kuitenkaan vielä ole merkittäviä. Tässä luvussa hahmottelemme organisaation näkökulmasta osallistumista avoimen datan ekosysteemiin. Liikkeelle kannattaa lähteä kevyesti, mutta tavoitellen näkyviä muutoksia ja todellista käyttöä datalle. Kun omat tietovarannot on identifioitu, voidaan niistä poimia helpoimmat avattavaksi ensin. Näiden alhaalla roikkuvien hedelmien kanssa päästään keräämään kokemuksia datan avaamisesta. Ensimmäisten pilottien perusteella voidaan organisaatiolle muodostaa tietovarantojen avoimuutta edistävä strategia, jonka mukaisesti rekistereitä ja rajapintoja järjestelmällisesti avataan.

Tässä luvussa esiteltävät datan avoimuudet mittarit (2.2) ja viitteellinen datavarantojen inventaario- ja julkaisuprosessi (2.3) muodostavat rungon seuraaville luvuille, joissa syvennetään maksuttomuuden, laillisen ja teknisen uudelleenkäytettävyyden, sekä datan löydettävyyden teemoja. Datan avaamishankkeessa on hyvä olla mukana laaja edustus organisaation osaamista, mukana tulisi olla ainakin informaatio-, laki-, talous-, tietotekniikka-, viestintä ja hallinto-osaamista sekä lisäksi päätösvaltaisia ihmisiä.

2.1 Datavarantoja toimijoita ja rooleja

Potentiaalisia toimijoita avoimen datan ekosysteemissä on niin julkishallinnon organisaatioissa, yrityksissä, järjestöissä kuin kansalaisten parissa. Julkishallinnolla on merkittävä rooli datan tuotannossa ja siksi olisi oleellista, että mahdollisimman suuri osa julkishallinnosta on mukana. Julkishallinnon datasta puhuttaessa tulee yleensä ensimmäisenä mieleen tunnetut perusrekisterit, kuten väestörekisteri, kiinteistörekisteri jne. sekä kenties tilastotiedot ja säätiedot. Myös kartta-aineistot ja paikkatieto on melko laajasti tunnettua ja jotkut muistavat mainita erilaiset liikennetiedot. Hallinnon data-aineistoja ei Suomessa ole kattavasti kartoitettu, mutta niiden määrä on eri arvioiden mukaan kuitenkin hämmästyttävän suuri.

Case: Dataa on kaikkialla – puurekisteri ja kirjaston tietokannat

Käytännön sovellusesimerkki New Yorkista, ja ajatuskulku kotimaisiin kirjastoihin liittyen osoittavat, että julkista dataa on kaikkialla kunhan sen olemassaolon huomaa.

1) Brett Camperin toteuttama Puut lähelläsi (Trees Near You) -palvelu sai kunniamaininnan New Yorkin kaupungin avoimen datan hyödyntämiskilpailussa 2010. Ilmainen iPhone-sovellus kertoo käyttäjälle sijainnin perusteella tietoja yli 500 000 elävästä puusta New Yorkin kaupungissa. Sovellus yhdistelee puhelimesta saatavaa GPS-sijaintitietoa, Wikipedia-artikkeleita puulajeista ja kaupungin avoimesti julkaisemaa puurekisteriä. Tämä on hyvä esimerkki siitä, kuinka paljon potentiaalisesti hyödyllistä dataa on olemassa mutta jonka olemassaoloa ei yleensä tule ajatelleeksikaan.

2) Kirjastojen tietokannoista voisi mainita aineistojen hyvät bibliografiset tiedot: tekijä, teoksen nimi, teoksen kansikuva, sisällysluettelo ja sisällönkuvailu. Useilla maakuntakirjastoilla on myös aluetietokantoja, esim. Tampereen maakuntakirjasto ylläpitää Pirkitta-aluetietokantaa, johon on tallennettu muun muassa kotimaisten aikakauslehtien artikkeliviitteitä vuodesta 1983 lähtien ja digitoituja näköispainoksia pienpainatteista. Tällaiselle paikalliselle datalle voisi löytyä soveltamisalueita, jos tietokanta sen sallisi.

Luonnollisesti jokaisesta lainatusta kirjasta jää myös jälki tietokantaan. Nyt tietokannoissa on mukana lainaajia koskevaa tietosuojan alaista dataa, mutta ilman henkilötietoja (anonymisoidut) ovat lainaustilastot, kuten lainatuimmat aineistot jne., olisivat mielenkiintoisia. Asiakkaalle itselleen olisi mielenkiintoista saada oma lainaushistoriansa käyttöön, hän voisi linkittää sen vaikka verkossa toimivaan lukupiirisovellukseen, jolloin hän pääsisi vaihtamaan kommentteja muiden ihmisten kanssa, jotka ovat lukeneet samoja kirjoja.

Julkishallinnon datavarantoja voidaan jäsentää esimerkiksi tuottajaorganisaatioiden, aineistojen sisällön tai datan oletettujen käyttötarkoitusten perusteella. Aiemmissa keskustelunavauksissa tarkastelua on kentän laajuuden takia yleensä rajattu esimerkiksi vain joihinkin seuraavista: paikkatieto (MMM 2005), organisatorisesti, esim. valtion budjettitalouden piiriin kuuluvien organisaatioiden tuottama data (Kuronen 1998). Tässä oppaassa laajennamme tarkastelua koskemaan periaatteessa kaikkea lain mukaisesti julkista, hallinnon tuottamaa tietoa.

On olennaista, että organisaatioissa tunnistetaan omat tietovarannot ja niistä jaetaan tietoa ja viestitään. Valtion tasolla on meneillään tietoarkkitehtuurin määrittelytyö (VM 2009), jonka puitteissa on tehty alustava kaavio suomalaisten datavarantojen kokonaiskenttää: «Tietomaaston kartoitusta – merkittäviä tietokokonaisuuksia» (Liite 3).

Vaikka oppaassa keskitytään nimenomaan julkishallintoon merkittävänä informaation tuottajana, niin ekosysteemin näkökulmasta ei ole olennaista, kuka tuottaa ja kuka käyttää informaatiota. Nykyään tilanne saattaakin olla jo niin,

että yksityiset yritykset tuottavat lähes yhtä paljon dataa kuin julkishallinnon organisaatiot. Enenevässä määrin myös yksittäiset kansalaiset osallistuvat datan tuotantoon ja sen jalostamiseen informaatioksi ja tiedoksi. Avoimen datan ekosysteemissä hallinnon organisaatiot, kansalaiset ja yritykset ovat kaikki paitsi tiedon käyttäjiä myös tiedon tuottajia – usein molempia samaan aikaan. Ekosysteemi-ajattelussa toimijoita ovat paitsi organisaatiot, niin myös yksittäiset ihmiset, jotka ovat keskenään vuorovaikutuksessa ja muokkaavat ekosysteemin käytäntöjä.

Moniin hallinnon prosesseihin, kuten esimerkiksi lakien ja asetusten valmisteluun, liittyy paljon organisaatorajat ylittävää viestintää ja tiedonvaihtoa. Lainvalmistelu saattaa esimerkiksi pitää sisällään Tilastokeskuksen tietojen hyödyntämistä ja budjettivaikutusten arvioimista valtionvarainministeriössä. Käytännöllisenä lähtökohtana voidaan pitää, että jokainen julkishallinnon organisaatio todennäköisesti sekä a) tuottaa uutta dataa että b) jalostaa, c) käsittelee ja d) hyödyntää jonkun muun tuottamaa dataa. Vaikka toimijakenttä on moninainen, eikä ole selkeästi osoitettavissa julkisen datan tuottajaorganisaatioita, niin toimijoiden roolit suhteessa dataan on kuvattavissa tarkasti (taulukko 2.1).

Taulukko 2.1. Esimerkkejä toimijoiden rooleista suhteessa dataan.

Tallentaja	raakadatan kerääminen ja tallentaminen
Jalostaja	raakadatan käsittely ja muokkaaminen
Aggregaattori	datan yhdistely ja koostaminen eri lähteistä
Harmonisoija	eri lähteistä tulevien tietojen yhdenmukaistaminen ja yhteismitallistaminen (samalta näyttävä asia myös tarkoittaa samaa)
Päivittäjä	tietojen päivittäminen
Julkaisija	datan julkaiseminen
Rekisterinpitäjä	datavarannon ylläpitovastuu
Sovelluskehittäjä datan loppukäyttäjänä	datan hyödyntäminen osana palvelua
Tulkitsija datan loppukäyttäjänä	datan tulkitseminen, esim. tutkija, yritys tai demokratiaaktivisti
Dataan pohjautuvien palveluiden (ks. luku 2) käyttäjät	ihmiset yritykset ja organisaatiot, jotka käyttävät avoimen datan päälle tehtyjä sovelluksia ja tulkintoja

Myöhemmin tässä luvussa esitettävän organisaation tietovarantojen inventaari-
on yhteydessä identifioidaan aineistot, joita organisaatiolla on hallussaan. Yllä-
mainittu roolijaottelu sekä liitteen 3. julkishallinnon tietomaaston kartoitus on
tarkoitettu avuksi inventaariin.

2.2 Datan avoimuuden mittareita

Keskustelunavausten, raporttien ja hankkeiden myötä avoimuudesta on nopeasti tullut keskeinen sana hallinnon sanastoon myös kun puhutaan tietojärjestelmistä ja tietovarannoista. Avoimuuskeskustelussa puhutaan samaan aikaan avoimesta lisensoinnista, teknisistä rajapinnoista, formaateista, metadatasta, tiedon harmonisoinnista, hallinnon läpinäkyvyydestä, ihmisten osallistumismahdollisuuksista, julkisuusperiaatteesta, uudelleenkäytettävyydestä ja kone-luettavuudesta. Ei ole aivan poikkeuksellista, että eri osapuolet tarkoittavat eri asioita. Aiheen laajuuden ja nopean kehityksen vuoksi väärinymmärrykset ja liika yleistäminen ovat yleisiä. Mitä oikeastaan tarkoitetaan avoimella julkishallinnon datalla?

Lainsäädännön mukaan julkisella datalla tarkoitetaan sellaisia aineistoja, jotka eivät kuulu yksityisyydensuojan piiriin – kuten henkilökisterit – tai ole muusta syystä laissa arkaluontoisiksi tulkittavia, kuten kansalliseen turvallisuuteen liittyvät aineistot.

Aineiston avoimuus ei ole mustavalkoinen joko-tai-kysymys, vaan siihen vaikuttaa ainakin aineiston tekninen saavutettavuus, löydettävyys, maksullisuus ja lisensointiehdot. Kaikkien aineistojen kohdalla äärimmäinen avoimuus ei ole edes tavoiteltavaa. Monesti aineiston uudelleenkäytettävyyttä voidaan merkittävästi parantaa ilman haittoja tekemällä ratkaisuja, jotka tavalla tai toisella lisäävät avoimuutta (esimerkiksi muuttamalla lisensointitietoja sallivammiksi tai tarjoamalla aineistot täysin veloitusetta).

Data-aineiston avoimuutta voidaan arvioida vertaamalla sitä avoimen datan kriteeristöön. Seuraava määritelmä pohjautuu Open Knowledge Definition -määritelmään (OKD). Julkisen datan täysin avoimella löydettävyydellä, saatavuudella ja uudelleenkäytettävyydellä tarkoitetaan seuraavaa:

Julkisella datalla tarkoitetaan aineistoja, joita eivät koske yksityisyydensuoja-, turvallisuus- tai muut rajoitteet, jotka tekisivät datasta salassa pidettävää tai arkaluontoista.

- Aineisto on löydettävissä ja kokonaisuutena saatavilla siten, että se voidaan ottaa laillisesti käyttöön välittömästi ilman viranomaisen avustusta esimerkiksi kello kolme yöllä.
- Aineisto on vapaasti tarjolla internetissä. Kuka tahansa saa lukea, ladata sen omalle koneelleen, kopioida, muokata, jakaa, tulostaa, etsiä, linkittää dataan, indeksoida, antaa data tietokoneohjelman syötteenä tai käyttää sitä mihin tahansa muuhun lainmukaiseen toimintaan ilman taloudellisia, juridisia, teknisiä tai käytännöllisiä rajoitteita.
- Käyttöehdoilla turvataan aineiston käyttäjälle varmuus aineiston alkuperästä ja aineiston tuottajalle oikeus tulla asianmukaisesti nimetyksi sen hyödyntämisen yhteydessä. Muita aineiston uudelleenkäyttöä ja jakelua rajoittavia ehtoja ei ole ja aineisto on saatavilla ilman rekisteröitymisvaatimuksia.
- Aineiston yhteyteen (esim. metatietoihin) sisällytetään tiedot aineiston alkuperästä ja käyttöehdoissa kerrotaan selkeästi aineiston vastaanottajalle, että hänellä on laillinen oikeus ja häntä jopa kannustetaan uudelleenkäyttämään aineistoa parhaaksi katsomallaan tavalla.

Koska täysin avoimen datan määritelmä harvoin kaikilta osin täyttyy tai on edes tarpeen täyttyä, on määritelmä purettu osiin: löydettävyys (2.2.1), kokonaisuus (2.2.2), käyttöehtojen tasa-arvoisuus (2.2.3), alkuperäisyys ja ajantasaisuus (2.2.4), laillinen ja vapaa uudelleenkäytettävyyys (2.2.5), maksuttomuus (2.2.6), kone-luettavuus (2.2.7), formaatin avoimuus (2.2.8) ja ymmärrettävyys (2.2.9). Tämä osiin purettu avoimen datan määritelmä toimii akselistona, jonka avulla voidaan arvioida yksittäisen julkisen datalähteen avoimuutta ja tunnistaa missä asioissa sitä kenties voitaisiin helposti lisätä.

2.2.1 Löydettävyys

Avoim: Aineiston olemassaolo ja sijainti on yleisesti tunnettu. Aineisto ja sen uudelleenkäytön sallivat lisensiointiehdot ovat helposti löydettävissä internetistä sekä ihmisille että hakuroboteille.

Suljettu: Aineisto on vain viraston operatiivisessa järjestelmässä, eikä sen olemassaolosta ole ulkopuolisilla lainkaan tietoa.

Google Maps -rajapinta tai Wikipedian sisältö ovat esimerkkejä aineistoista, joiden olemassaolo ja laillinen käytettävyys on yleisesti tiedossa. Aineiston löydettävyyttä voidaan parantaa lisäämällä se hyvin ylläpidettyyn datakatalogiin, optimoimalla datavarantojen kuvailutiedot hakukoneita varten ja julkaisemalla data linkitetyn datan paradigman mukaisesti. Yleistä tunnettuutta voidaan vahvistaa tiedottamalla aktiivisesti potentiaalisille uudelleenkäyttäjille datavaran olemassaolosta verkossa, julkaisuissa ja tapahtumien yhteydessä. (ks. luku 6: Avoimen datan infrastruktuuri)

2.2.2 Kokonaisuus

Avoim: Data on kokonaisuudessaan vapaasti ladattavissa internetistä. Saavutettavuutta ja datan käyttömahdollisuuksia ei rajoiteta epäsuorasti, tarjoamalla pääsy vain osaan tietokannasta kerrallaan.

Suljettu: Aineistosta on saatavilla avoimesti vain osa ja koko aineiston saaminen edellyttää erillistä sopimusta.

Tyypillinen tapaus, jossa aineiston kokonaisuutta rajoitetaan tarkoituksella tai tahattomasti on sellainen, jossa aineisto tarjotaan vain kyselyrajapinnan kautta, eikä aineistoa ole mahdollista ladata kokonaan. Mikäli aineisto on saatavilla kokonaisuudessaan tarkoittaa se, että kenen tahansa on teknisesti mahdollista myös ryhtyä jakamaan dataa itseään ja muita varten. Kokonaisuuden rajoittaminen voi olla keino varmistaa aineiston loukkaamattomuus (integriteetti) ja estää rinnakkaisten kopioiden syntyminen aineistosta. Toisaalta kokonaisuuden rajoittaminen estää datan kattavaan analyysiin perustuvan käytön ja aiheuttaa kyselyrajapinnan kuormitusta, joka voitaisiin välttää tarjoamalla oma kopio. (ks. luku 5: Tekninen valmistautuminen)

2.2.3 Käyttöehtojen tasa-arvoisuus

Avoim: Aineisto on saatavilla kaikille ja kaikkiin laillisiin käyttötarkoituksiin yhtä helposti. Datan luovutuksessa ei aseteta käyttäjiä eikä käyttötarkoituksia eriarvoiseen asemaan, vaan muut hallinnon toimijat, kansalaiset, yritykset ja myös ulkomaalaiset toimijat ovat tasa-arvoisia.

Suljettu: Aineiston tarjoamista on rajoitettu käyttäjän tai käyttötarkoituksen mukaan, esimerkiksi vain tutkimus ja tuotekehityskäyttöön tai ainoastaan epäkaupalliseen käyttöön tai ainoastaan suomalaisille.

Käytännössä tasa-arvoisuus toteutuu, mikäli aineisto on saatavissa verkosta ilman rekisteröintivaatimuksia, jolloin kuka tahansa voi ottaa sen käyttöönsä standardinmukaisilla lisenssiehdoilla. Lisenssi ei estä ketään käyttämästä dataa jollakin määrättyllä käyttöalueella. Erityisesti myös kaupallinen käyttö on sallittua, sillä kaupallisten toimijoiden toivotaan liittyvän avoimen datan ekosys-

teemiin. Tasa-arvoisuus tarkoittaa ennakoivasta kontrollista luopumista. Dataa saa käyttää myös taitamattomasti ja poliittisiin tarkoituksiin. (ks. luku 3: Lupa julkaista ja käyttää)

2.2.4 Alkuperäisyys ja ajantasaisuus

Avoin: Aineisto on tarjolla alkuperäisessä muodossaan ja alkuperäisellä tarkkuusasteella silloin, kun se ei uhkaa yksilöiden tietosuojaa. Aineisto julkaistaan lähimpänä sen syntypaikkaa (alkuperäisen tuottajan toimesta) ajantasaisesti suhteutettuna aineiston sisällön muuttumisnopeuteen.

Suljettu: Aineisto julkaistaan viivästetysti, alkuperäistä pienemmällä tarkkuudella tai ainoastaan yhdistetyssä muodossa muiden aineistojen kanssa.

Raakadatan lisäksi myös yhdistettyjä ja jalostettuja muotoja datasta voidaan asettaa vapaasti saataville. Joissain tapauksissa yleistämällä ja laskemalla tarkkuustasoa voidaan muutoin yksityisyydensuojan kannalta arkaluontoinen aineisto saattaa julkiseksi. Anonymisointi ja yleistäminen täytyy tehdä kuitenkin erityisellä huolella. (ks. luku 5: Tekninen valmistautuminen)

2.2.5 Laillinen ja vapaa uudelleenkäytettävyys

Avoin: On olemassa standardikäyttöehdot, joiden puitteissa aineisto on vapaasti uudelleenkäytettävissä kaikille riippumatta käyttäjästä tai käyttötarkoituksesta. Mahdollisia yksityisyydensuojan, turvallisuuteen liittyviä rajoituksia lukuunottamatta datan käytölle ei ole asetettu muita juridisia rajoituksia. Lisensointiehdot on esitetty selvästi ja läpinäkyvästi siten, että ne kannustavat uudelleenkäyttöön. Aineistoon mahdollisesti kohdistuvista tekijänoikeuksista on luovuttu (Waiver of Rights) ja se kerrottu suoraan aineiston uudelleenkäyttöön kannustavissa käyttöehdoissa.

Suljettu: Aineisto on lisensoitu uudelleenkäyttöä rajoittavalla lisenssillä, siihen kohdistuu tekijänoikeuksia, tai sen uudelleenkäytön sallimista ei ole suoraan esitetty missään.

Salliva lisensointi voidaan toteuttaa esimerkiksi Creative Commons, Open Database -lisenssillä tai vastaavalla. Aineistoon mahdollisesti kohdistuvista tekijänoikeuksista tulisi avoimen julkisen datan osalta luopua, esimerkiksi Creative Commons nolla -lisenssillä, jolloin vältetään epäselvyyksiltä koko jatkojalostusketjussa. Yleisin haastatteluissa esiin noussut toive käyttöehtoihin liittyen on halu tietää ketkä aineistoa käyttävät ja mihin. Hyvin usein ei ollut mitään varsinaisia tarpeita rajoittaa käyttötarkoituksia, mutta ne haluttiin kuitenkin tietää oman toiminnan kehittämisen näkökulmasta. Aineistojen käytön seuranta voidaan tehdä myös ilman allekirjoitettavia sopimuksia ja rajoittavia käyttöehtoja, esimerkiksi verkossa tapahtuvan käyttäjärekisteröinnin avulla. (ks. luku 3: Lupa julkaista ja käyttää)

2.2.6 Maksuttomuus

Avoin: Data on saatavissa täysin maksutta.

Suljettu: Dataa tarjotaan maksuperustaisesti ja myynnistä saaduilla tuotoilla katetaan dataa tuottavan organisaation muita toimintakuluja.

Enintään irtiottokustannuksilla tarjottavaa aineistoa voidaan pitää melko avoimena. Hyvin usein kuitenkin aineiston ylläpidon ja tuotannon kustannukset ovat moninkertaiset suhteutettuna irtiottokuluina perittäviin maksuihin. Pienikin maksu rajoittaa aineiston käyttöä merkittävästi siihen liittyvän vaivalloisuuden ja sopimusten takia. Voi olla, että suurin osa irtiottokustannuksista muodostuu itseasiassa sen laskuttamiseen liittyvästä byrokratiasta, jolloin irtiottokustannuksien perimistä ei voida pitää perusteltuna. Mikäli erityisestä syystä peritään maksuja, tulisi maksu olla mahdollista suorittaa internetissä ja saada aineisto käyttöön välittömästi ilman viranomaisen työtä. (ks. luku 5: Avoimen datan taloudesta)

2.2.7 Koneluettavuus

Avoim: Datavarannolla on pysyvä sijainti internetissä ja sitä pääsee lukemaan automatisoidusti ja ohjelmallisesti. Data on riittävässä määrin strukturoitu automatisoidun prosessoinnin mahdollistamiseksi. Myöskin lisensointiehdot ovat koneluettavia, ne voidaan hyväksyä verkossa ja aineisto saadaan käyttöön ilman viranomaisen työtä.

Suljettu: Data julkaistaan täysin epärakenteellisessa muodossa siten, että sen tulkinta on mahdollista vain ihmislukijalle, esim. pdf-dokumentit tai html-sivut.

Peukalosääntönä koneluettavuudelle voidaan pitää sitä, että etevä ohjelmoija kykenee verrattain lyhyessä ajassa tekemään ohjelman, joka automaattisesti hakee datan internetistä, lukee sen ohjelman muistiin ja tulostaa uudelleen muokattuna vaikkapa iPhoneen näytölle. Mikäli aineisto ei missään vaiheessa ole koneluettavassa muodossa, on sen saattaminen sellaiseksi varsin työlästä. Usein kuitenkin esimerkiksi organisaation verkkosivuilla tarjotaan aineistoja ei-koneluettavasti HTML-muodossa, mutta samat aineistot ovat jossain toisessa järjestelmässä myös koneluettavassa muodossa, jolloin niiden julkaiseminen on jo paljon helpompaa. (ks. luku 5: Tekninen valmistautuminen)

2.2.8 Formaatin avoimuus

Avoim: Aineisto on saatavilla käyttökelpoisessa ja muokattavassa formaatissa, jonka määrittely ja kehitys eivät ole minkään yksittäisen yrityksen tai vastaavan hallussa.

Suljettu: Data on tarjolla ainoastaan formaatissa, joka on jonkin yrityksen hallinnoima ja jonka hyödyntämiseksi on pakko käyttää saman yrityksen ohjelmistoja.

Tämä voidaan saavuttaa tarjoamalla aineisto avoimessa formaatissa, esimerkiksi sellaisessa, jonka spesifikaatio on julkisesti ja vapaasti saatavilla ja joka ei aseta rahallisia tai muita rajoitteita formaatin käytölle. Samaa dataa kannattaa mahdollisuuksien mukaan tarjota useammassakin formaateissa. Aina avoimen formaatin käyttöön ei ole realistisia mahdollisuuksia. Esimerkiksi jotkut paikatieto-järjestelmät käyttävät valmistajakohtaisia formaatteja, jolloin avoimiin formaatteihin voidaan siirtyä vain järjestelmäuudestuksen yhteydessä. (ks. luku 5: Tekninen valmistautuminen)

2.2.9 Ymmärrettävyys

Avoin: Data on kuvailtu ja dokumentoitu kattavasti ja selkeästi.

Suljettu: Aineisto on tarjottu formaatissa, jonka käyttöä tai sisältöä ei ole kuvailtu uudelleenkäytön edellyttämällä tavalla.

Aineiston ymmärrettävyyttä ja sitä kautta uudelleenkäytettävyyttä voidaan merkittävästi parantaa metadatan, dokumentaation, käyttöesimerkkien ja laatumääritelmien avulla. Hyvän dokumentaation ainoa haittapuoli on sen tekemisen työläys. Ohjeistaminen saattaa muodostua merkittäväksi aineistojen julkaisemisen hidasteeksi. Toisaalta dokumentaatio voi joskus olla hyvinkin kevyt ja sitä voi parantaa myöhemmin. Esimerkiksi sarkainerotellussa tiedostossa riittää, että sarakeotsikot ovat tiedostossa mukana. (ks. luku 5: Tekninen valmistautuminen)

2.3 Datavarantojen inventaario- ja julkaisuprosessi

Hyvin kuvailtu laadukas data luotettavasti toimivien rajapintojen kautta jaeluna johtaa avoimen datan suurimpiin hyötyihin. Datan avaaminen vaatii kuitenkin aikaa ja resursseja, eikä kaikkea voida rakentaa kerralla valmiiksi. Datan jakamisessa kannattaa etsiä pieniä toimia, joilla datan avoimuutta voidaan merkittävästi lisätä.

Esitämme tässä datan avaamiseen vaiheittaisen ja käytännön kokeilujen kautta ohjautuvan etenemismallin. Tavoitteena on, että datan avoimesta jakelusta ja ekosysteemiin osallistumisesta saadaan kokemuksia aiemmin, kuin mitä tietojärjestelmien uusimissykli toisi mukanaan. Nähdään mille datalle on eniten kysyntää ja saadaan palautetta siitä, missä formaateissa ja millaisen rajapinnan kautta mitäkin kannattaa jakaa. Prosessi on tarkoitettu viitteelliseksi ja ennen kaikkea herättämään ajatuksia.

Datan avaamisessa kannattaa välttää turhia ennakko-oletuksia datan käyttötarkoituksesta ja potentiaalisista hyödyntäjistä. Joillekin aineistoille voi olla piilevää kysyntää, joka paljastuu vasta kun aineisto saa julkisuutta. Esimerkiksi paikkatietolainsäädännöstä käyty keskustelu on lisännyt paikkatietoaineistojen tunnettuutta ja sitä kautta niiden kysyntää. Priorisointimielessä on loogista selvittää mille tietovarannoille on eniten kysyntää. Mielikuvien sijaan on kuitenkin parempi nojautua kokeiluihin, kannustaa datan hyödyntämiseen ja seurata mitä tapahtuu. Yksikin julkisuutta saanut sovellus saattaa johtaa aineiston kysynnän lisääntymiseen ja muuttaa tilannetta lyhyessä ajassa merkittävästi.

Teknisesti tietovarantojen julkaisu liittyy organisaation tietojärjestelmien ja etenkin tietoarkkitehtuurin kehittämiseen. Aineistojen julkaisua voidaan ajatella tietoarkkitehtuurin kehittämisen tukitoimintona eikä niinkään arkkitehtuuriprojektin lopputuloksena. Sen sijaan, että luvataan avoimet rajapinnat seuraavan järjestelmäuudistuksen yhteydessä, lähdetäänkin siitä, että ensin julkaistaan organisaation tietovarantojen inventaarion tulokset. Sen jälkeen julkaistaan joitain kiinnostavia tietoaineistoja ladattavaksi sellaisena, kuin ne ovat. Datan julkaisusta saadut kokemukset voivat toimia tärkeänä syötteenä niin aineistojen harmonisointiin kuin yleisen tietojärjestelmien kehitystyöhönkin.

Vaiheittaisen prosessin (kuva 2.1) jokaisella kierroksella syvennetään organisaation omien tietoresurssien tuntemusta, julkaistaan uutta materiaalia ekosysteemin käytettäväksi ja kerrytetään organisaation kokemuksia avoimeen dataan

liittyvästä toiminnasta. Prosessin osia ovat: **a) aineistojen analysointi**, **b) aineiston julkaiseminen** ja **c) datan avaamisesta oppiminen**.

Ensimmäinen kierros: Datasta ilmoittaminen

Datan avaamisesta ja organisaation hallussa olevista aineistoista ilmoittaminen pyritään tekemään mahdollisimman nopeasti ja kevyesti. Tässä vaiheessa vastataan kysymyksiin:

- 1) Mitä aineistoja meillä ylipäättään on?
- 2) Mitkä niistä ovat julkisia?
- 3) Kuinka avoimia eri aineistot nykyisin ovat? (ks. mittareita datan avoimuudesta)

Toinen kierros: Pilottiaineistojen julkaiseminen

Kun omat tietovarannot on tunnistettu, niistä voidaan poimia helpoimmat avattaviksi ensin. Seuraavalla iteraatiokierroksella pilotoidaan teknisesti, juridisesti ja taloudellisesti helpoimmin avattavien aineistojen julkaisua. Tässä vaiheessa vastataan kysymyksiin:

- 4) Mitkä aineistot ovat teknisesti helposti avattavissa?
- 5) Minkälaisin käyttöehdoin niitä halutaan julkaista?
- 6) Ketkä organisaatiossa vastaavat eri aineistojen sisällöstä ja ketkä teknisistä järjestelmistä?

Kolmas kierros: Dokumentointi ja käyttötapaukset

Julkaistujen aineistojen määrä voidaan kierros kierrokselta lisätä, mutta samalla on hyvä lisätä ja selkiyttää aineistojen dokumentaatiota uudelleenkäyttäjiä varten sekä koota aineistojen käyttötapauksia ja muuta tietoa organisaation sisältä. Tässä vaiheessa vastataan kysymyksiin:

- 7) Miten tieto on kuvailtu sisällöllisesti ja dokumentoitu teknisesti?
- 8) Millaisia tiedontarpeita, käytäjäryhmiä ja käyttötapauksia eri aineistojen nykyiseen käyttöön liittyy?
- 9) Minkälaisia tiedon saatavuuteen ja käytettävyyteen liittyviä pyyntöjä ja tarpeita on noussut esiin organisaation sisältä ja ulkopuolelta?

Neljäs kierros: Tietoarkkitehtuuri ja sanastotyö

Ensimmäisten pilottien kokemuksella voidaan organisaatiolle muodostaa tietovarantojen avoimuutta edistävä strategia, jonka mukaisesti rekistereitä ja rajapintoja järjestelmällisesti avataan. Sen jälkeen on mahdollista kehittää tietovarantojen käyttöä myös yli organisaatorajojen. Tässä vaiheessa vastataan kysymyksiin:

- 10) Minkälaisia organisaatorajat ylittäviä tiedon harmonisointi- ja yhdenmukaistamistarpeita ja mahdollisuuksia on?
- 11) Minkälaisiin järjestelmä uudistuksiin tietovarantojen avoimuuden parantaminen liittyy?
- 12) Miten organisaation käyttämiä datavarantoja voitaisiin paremmin organisoida yhdessä muiden toimijoiden kanssa?

Kuten jo aikaisemmin mainittiin, niin prosessi on viitteellinen ja toteutus saattaa vaihdella organisaatiottain. Lähtöpointe on kuitenkin selvä. Datan avaaminen lähtee aina omien datavarantojen tunnistamisesta ja nykytilan arvioimisesta. Tämän jälkeen on mahdollista osallistua avoimen datan ekosysteemiin – ensin projektimielessä kokeillen ja sen jälkeen sisällyttäen avoin data osaksi organisaation pysyvää toimintatapaa. Liikkeelle voidaan lähteä pienestä ja kertyneet kokemukset antavat aineksia organisaation omaan avoimen datan strategiaan.



Kuva 2.1 Datan avaamisessa voidaan edetä vaiheittain vuorovaikutuksessa sen hyödyntäjien kanssa.

2.3.1 Datasta ilmoittaminen

Organisaation edustaja kysyy datan uudelleenkäyttöyhteisöiltä: «Kertokaa meille, mitä tietoa haluatte ja missä muodossa, niin katsomme, mitä voimme asian hyväksi tehdä». Uudelleenkäyttäjät vastaavat: «Kertokaa meille, mitä dataa teillä on, niin kerromme, mitä haluamme»

Aivan ensimmäinen julkaistava data tulisi olla listaus aineistoista, joita organisaatiolla ylipäätään on olemassa, vaikka nämä aineistot eivät vielä olisi saatavilla. Tällaista organisaation nettisivuilla tai laajemmassa useiden organisaatioiden tietoja yhdistävässä datakatalogissa julkaistavaa listausta kutsutaan organisaation julkisten tietoaaineistojen rekisteriksi (Information Asset Registry IAR). Listaukseen sisältyvät sekä jo avoimet datavarannot että tietolähteet, jotka eivät

vielä ole avoimesti saatavilla, mutta ovat kuitenkin julkisia ja joiden avoimuutta voidaan jatkossa lisätä.

Monissa organisaatioissa tällainen listaus omista tietovarannoista saattaa olla jo olemassa, toisissa sitä varten pitää tehdä omien datavarantojen inventaario. Yksi ensimmäisistä tehtävistä datan avaamista käynnistettäessä on perata olemassaolevat tietovarannot ja niihin liittyvät käytännöt.

Ensimmäinen lista organisaation datavarannoista syntyy yksinkertaisimmin muistinvaraisesti ja omia verkkosivuja selaamalla. Kattavampi listaus saadaan, jos kartoitustyöhön osallistuu useampia henkilöitä. Esimerkiksi muutaman tunnin työpajassa pystytään tuottamaan julkaisukelpoinen listaus, joka sisältää olennaisimmat tiedot organisaation hallussa olevista aineistoista. Tavoitteena ei ole vielä kattavasti selvittää kaikkia tietoja kaikista aineistoista. Alla on muutamia tärkeimpiä kysymyksistä, joihin pyritään aineistokohtaisesti vastaamaan.

- Aineiston nimi
- Lyhyt kuvaus aineiston sisällöstä
- Kuinka avointa aineisto on nykyisin, onko se saatavilla käyttöön ja jos on niin miten?

Kun julkisten tietovarantojen rekisteri on laitettu avoimesti verkkoon, voidaan siitä jo viestiä potentiaalisille datan uudelleenkäyttäjille, luoda kontakteja ja kerätä mielipiteitä. Organisaatio voi esimerkiksi jalkautua kehittäjäyhteisöihin sosiaalisen median ja tapahtumien kautta. Viestinnässä kannattaa kiinnittää huomiota kannustavuuteen ja siihen, millaisia odotuksia herättää yleisössä. Jos aineistojen avaamisesta ei ole vielä tehty päätöksiä ja halutaan pelkästään ideoita niiden hyödyntämisestä, kannattaa se kertoa.

Voi olla, ettei käyttäjien huomiota aivan alussa saavuteta. Parhaimmillaan kuitenkin jo pelkän omien datavarantojen listauksen avulla voidaan käynnistää organisaation ja datan hyödyntäjien välinen dialogi ja saada palautetta siitä, mitä dataa kannattaisi julkaista mitenkään. Lopuksi palaute ja datan avaamisesta saadut kokemukset kootaan ja arvioidaan. Mikä toimi, mikä ei? Valmistaudutaan siirtymään seuraavalle iteraatiokierrokselle, jossa pilotoidaan datan julkaisemista.

2.3.2 Pilottiaineistojen julkaiseminen

Kun organisaation keskeisimmät tietovarannot on tunnistettu, on tärkeää erotella huolellisesti julkaisemiseen kelpaavat aineistot muista. Tietokannoissa saattaa olla sekaisin julkisia ja salaisia tietoja. Esimerkiksi monet valtion työntekijöiden nimet ovat periaatteessa julkisia, mutta esimerkiksi hävittäjälentäjistä ei haluta kansallisen turvallisuuden vuoksi kertoa mitään. Nopean datan avaamisen kannalta tällaiset kokonaisuudet kannattaa lykätä myöhemmäksi. Erotellaan aineistot täysin julkisiin, epäselviin ja ei-julkisiin. Aluksi voidaan keskittyä pelkästään selvästi julkisten aineistojen avaamiseen.

Vielä ei kannata siivota listalta pois sellaisia aineistoja, joiden julkaisemiseen organisaatiolla ei ole oikeutta. Mahdollisuuksien mukaan myös näistä aineistoista kannattaa kertoa. Organisaation ulkopuolisten toimijoiden kannalta tällainen tieto on erittäin arvokasta hallinnon tietovarantoihin liittyvän kokonaiskuvan hahmottamisen kannalta.

Seuraavaksi suunnitellaan, ketkä toimivat organisaation yhteyshenkilöinä datan hyödyntäjiin päin ja miten esiintuleviin asioihin reagoidaan. Voidaan esimerkiksi varautua etukäteen nostamaan esiin mielenkiintoisia datan käyttötapoja ja miettiä, kuinka konfliktitilanteissa toimitaan ja miettiä vastauksia, miten esimerkiksi reagoidaan datan epämoraaliseen käyttöön.

Itse datan jakamiseen ei aluksi tarvitse tehdä minkäänlaista käyttöliittymää.

Kaikkien kannalta on helpointa, jos datan hyödyntäjiä ei pyydetä rekisteröitymään tai tilaamaan aineistoa organisaatiolta, vaan julkisesti kerrotaan osoite, josta data on saatavilla. Raakadatan jakamisella pelkistetysti tarkoitetaan, että jos aineisto on tietokantamuodossa, niin tietokannan taulujen sisältö tulostetaan vaikka pilkkueroteltuun tiedostoon, joka laitetaan internet-palvelimelle. Jos aineisto on jo valmiiksi tiedostoissa, riittää, että nämä tiedostot laitetaan julkisesti saataville.

Raakadatan jakamisen tarkoituksena ei välttämättä ole se, että joku toteuttaisi suoraan järjestelmiä siihen perustuen. Avoimesti julkaistu data auttaa kuitenkin eri tahoja perehtymään tietosisältöön ja suunnittelemaan sen hyödyntämistä. Se tekee kyseisen datan olemassaolon tunnetuksi. Kun kiinnostus tietoa kohtaan on herännyt, dataa tarjoava organisaatio voi yhdessä hyödyntäjien kanssa suunnitella sen käyttöä parhaiten palvelevat palvelurajapinnat tai muut jakeluratkaisut.

Asiaa ei ole syytä mutkistaa ja datan avausta viivyttää esimerkiksi sen takia, että organisaation resurssit eivät juuri nyt riitä rajapintojen määrittelyyn tai viimeistä piirtoa myöten hienon dataportaalin rakentamiseen. Tässäkin iteratiivisuus on tärkeää ja pienin askelin päästään pidemmälle. On parempi laittaa aineisto saataville ensin kevyesti, mutta kuitenkin kokonaisuutena ja rakenteellisessa muodossa, ja vasta sitten alkaa miettiä mahdollisia rajapintoja. Pinnan kiillotus ja käyttöliittymät ovat tärkeysjärjestyksessä viimeisimpiä.

2.3.3 Dokumentointi ja käyttötilanteet

Viitteellisen datan avaamisprosessin mukaan on tehty jo kaksi täyttä kierrosta. Ensimmäisen tuloksena julkaistiin lista kaikista organisaation hallussa olevista julkisista aineistoista. Toisen tuloksena julkaistiin ensimmäiset helposti avattavat aineistot. Samalla organisaation omien tietoresurssien tuntemus on vahvistunut ja datan käyttäjäyhteisöihin on syntynyt ensimmäiset kontaktit. Tämän seuraavan dokumentointi ja käyttötilanteet-kierroksen aikana tehdään kattavampi datavarantojen inventaario, julkaistaan uusia aineistoja ja parannetaan niiden julkista dokumentaatiota.

Inventaarioprosessi datan avaamista varten on pitkälti samantyyppinen prosessi, kuin mitä organisaatioissa saatetaan tehdä muutoinkin pelkästään sisäiseen käyttöön omien tietovarantojen jäsentämiseksi ja uudelleenorganisoinniseksi. Inventaariossa tunnistetaan myös omia tietotarpeita organisaation työntekijöiden näkökulmasta. Tämä on pelkkää teknistä datan avaamista työläämpää, mutta toisaalta se tukee organisaation toimintaa laajemmin.

Analysoimalla käytössä olevia järjestelmiä tietohallinnon kanssa saadaan hyvä pohjakäsitys organisaation tietoresursseista ja niihin liittyvistä käytännöistä. Tietohallinnossa ei kuitenkaan välttämättä tunneta aivan kaikkia organisaatiossa käytettyjä aineistoja ja niiden käyttötapoja. Käsitystä voidaan syventää haastatteleamalla eri käyttäjäryhmiä ja kuvailemalla käyttötilanteita, joissa organisaatiossa tarvitaan tietoa. Käyttäjien näkökulma voi auttaa löytämään uusia tiedon hyödyntämismahdollisuuksia, joita voidaan nivoa osaksi datanavausprojektia.

Tämän iteraatiokierroksen julkaisuvaiheessa voidaan kehittää jo aiemmin julkaistujen aineistojen jakelua esimerkiksi toteuttamalla avoin rajapinta (ks. luku 5.5). Aineistojen sisällön ja käyttöoikeuksien kuvailu ja tekninen dokumentaatio ovat myös tärkeitä kehityskohteita.

2.3.4 Tietoarkkitehtuurin kehittäminen ja sanastotyö

[Pienestä datavarantojen listauksesta](#) liikkeelle lähtenyt prosessi voi ajan myötä kulminoitua tietoarkkitehtuurin kehittämiseen, järjestelmien suunnitteluun ja organisaatorajat ylittävään tietovarantojen yhtenäistämiseen. Olennaista on, että matkan varrella avattujen datavarantojen hyödyntämisestä saattaa jo olla kertynyt kokemuksia.

Kuten aiemminkin iteraatiokierroksilla, jälleen organisaatioiden omien tietoresurssien dokumentaatiota täydennetään, aineistojen julkaisua parannetaan ja datan avaamisen oppeja otetaan mukaan strategiseen työhön. Tavoitteena on mallintaa organisaatiossa tapahtuvaa tiedon käsittelyä ja tarpeita ja kehittää tietovarantojen yhteiskäyttöä myös yli organisaatorajojen. Tämän iteraatiokierroksen julkaisuvaiheessa organisaatiolla on laadukkaita tietoaaineistoja, jotka on harmonisoitu yhteiskäyttöisiksi muiden organisaatioiden kanssa ja julkaistu hyvin dokumentoidusti.

Verkkopalveluiden ja tietokantojen lisäksi kannattaa tässä vaiheessa kiinnittää huomiota myös dokumentteihin, joita käytetään toistuvasti - tällaisia voivat esimerkiksi olla taulukkolaskentatiedostot. Tiedon käyttämisen lisäksi pyritään löytämään kehittämiskohteita. Käyttäjiltä kysytään minkälaisiin tietoon liittyviin tarpeita heillä on, minkälaisiin ongelmiin he ovat törmänneet ja minkälaisia tarpeita arvelevat muilla käyttäjillä olevan. Tietolähteiden saatavuuden lisäksi selvitetään, onko tieto sellaisessa muodossa, että siitä on hyötyä vai liittyykö sen hyödyntämiseen jonkinlaista kitkaa tai mekaanista työtä.

Jos tietoarkkitehtuurista alkaa esimerkiksi löytyä dokumentteja ja kansioita, joilla ei ole selkeästi kohdennettavaa luojaa tai tekijää, tai ne on arkistoitu poikkeavasti, on tämä yleensä selkeä merkki esimerkiksi a) toimimattomasta käytännöstä, joka on ohitettu (kuten hankala toimintamalli), b) toimintatavasta, jota ei ole ohjeistettu riittävästi (ohjeiden puute) tai c) uudesta tarpeesta, jota ei ole ennakoitu (kuten uusi yksikkö, projektiryhmä etc.)

Useat nykyisistä tietoarkkitehtuurin haasteista kietoutuvat saman asian ympärille: tietoa tuotetaan erittäin runsaasti ja epäyhtenäisesti. Alkuperäisiä työvälineitä ei ole välttämättä suunniteltu aineistojen nykyisen paljouden hallintaan. Aineistojen yhdistäminen muodostuu keskeiseksi ongelmaksi organisaatioiden sisällä ja organisaatioiden välisessä yhteistoiminnassa. Huonossa tapauksessa lähtötilanteessa ei ole saatavilla kunnollista kuvausta siitä, mitä talletettua tietoa on missäkin järjestelmässä. Kukin toimija katsoo asiasta vain omasta näkökulmastaan, tiedot ovat hajallaan eri järjestelmissä ja tietoa samoista asioista kerätään ja päivitetään useissa paikoissa.

Aineistojen kuvaamisella ja menetelmien määrittelyllä pyritään ratkomaan ja ennaltaehkäisemään yhteensopivuusongelmia ja luomaan kattava näkemys aineistoihin yli organisaatorajojen. Organisaatioiden sisällä ja myös organisaatioiden välillä pyritään pikkuhiljaa yhdenmukaistamaan tietoarkkitehtuuria ja löytämään yhteisiä tapoja kuvata tietoa.

Tietoarkkitehtuurin laatiminen edellyttää eri osapuolten organisoitua yhteistoimintaa, aikaa ja runsasta työtä. Mitä laajempi näkemys pyritään saavuttamaan, sitä työläämpää se on. Esimerkki tällaisesta tietoarkkitehtuurihankkeesta on hiljattain valmistunut kaksivuotinen KuntaGML-hanke, jonka tavoitteena oli ottaa kunnissa käyttöön standardimuotoiset rajapinnat paikkatiedolle, kuten kaavoituksen pohjatiedoille. Hyvä arkkitehtuuri palkitsee työläydestään huolimatta järjestelmien yksinkertaisempaan ylläpitöön, säästyneenä työnä, tietovarantojen laadun paranemisena ja ennen kaikkea mahdollistaa tietojen laajemmän uudelleenkäytön.



**Epäselvät tai kokonaan puuttuvat
käyttöehdot rajoittavat usein aineistojen
käyttöä paljon enemmän kuin aineiston
haltija on tarkoittanut.**

Tässä luvussa käymme tiivistetysti läpi keskeisen lainsäädännön ja sääntelyn, joka vaikuttaa datavarantojen avoimeen ja maksuttomaan jakeluun verkossa. Tarkastelussa ovat datan maksullisuus, datan käyttöoikeudet, kilpailulainsäädäntö ja EU-sääntely. Käsiteltävät lait ja säännökset on listattu taulukossa 4.1.

Datavarantojen avaamissuunnitelmien yhteydessä on selvitettävä, mitkä organisaation hallitsemista aineistoista ovat julkisia. Tarkastelu on tehtävä ensisijaisesti tietosuojalainsäädännön ja viranomaisen toiminnan julkisuudesta säädetyn lain perusteella. Mikäli näiden lakien soveltaminen tuottaa ristiriitaisen lopputuloksen aineiston julkaisemiselle, ensisijainen tulkinta on tehtävä tietosuojasäännösten perusteella.

Monissa tapauksissa yksityisyydensuojan turvaaminen on jo ratkaistu muita tarkoituksia varten, esimerkiksi anonymisoimalla raakadataa tai esittämällä se karkeammalla tarkkuustasolla, minkä takia yksilöitä ei voida tunnistaa. Julkisen datan maksullisuuteen tai maksuttomuuteen vaikuttaa valtion maksuperustelaki ja joissain tapauksissa myös kilpailulainsäädäntö.

Kun tavoitteena on tietovarantojen käytön huomattava lisääminen, tulisi yrityksille, kansalaisille ja muille organisaatioille antaa lailliset oikeudet uudelleenkäyttää julkista dataa. Julkisuuden ja maksuttomuuden ohella tärkeä datan hyödynnettävyyteen vaikuttava seikka on aineistoihin mahdollisesti kohdistuvat tekijänoikeudet ja lisensointikäytännöt. Julkishallinnon organisaatiot voivat kannustaa datan uudelleenkäyttöön lisensoimalla tietovarannot sallivasti. Epäselvät tai kokonaan puuttuvat käyttöehdot ja hankala sopimusbyrokratia ovat tekijöitä, jotka usein käytännössä rajoittavat aineistojen käyttöä paljon enemmän kuin aineiston haltija on tarkoittanut. Selkeästi ilmoitetut ja standardeja noudattavat käyttöehdot vähentävät niin aineiston tuottajien kuin aineiston käyttäjienkin kuormitusta.

Avointa dataa omassa liiketoiminnassaan hyödyntävät yritykset tarvitsevat sopimuksia tai muita takeita tietolähteen tarjonnan jatkuvuudesta ja palvelutasosta. Tällaisten erillissopimusten teko ja korkean palvelutason takaaminen maksullisesti ei ole ristiriidassa sen kanssa, että raakadata tarjottaisiin sellaisena kuin se on saatavilla maksutta ja ilman sopimuksia. Tarkoitus ei ole ylläpitää keinotekoisesti kahta järjestelmää, joista toinen olisi luotettava ja toisen kautta data olisi heikompilaatuista. Esimerkiksi Googlen karttapalvelua voi käyttää pienillä volyymeilla ilmaiseksi, mutta kuormituksen kasvaessa tehdään sopimus Googlen kanssa.

Taulukko 3.1. Julkishallinnon avoimeen dataan liittyvää lainsäädäntöä

Julkisuuslaki	Julkisuuslailla tarkoitetaan vuonna 1999 voimatulleita lakia ja asetusta viranomaisten toiminnan julkisuudesta (1999/621) ja (1999/1030).
Tietosuojalainsäädäntö	Yleisnimityksellä tietosuojalainsäädäntö tarkoitetaan ainakin ensisijaisesti henkilötietolakia (523/1999) ja myös muita yksityisyyden suojaan liittyviä säädöksiä kuten rikoslain (24. luvun) pykälää yksityiselämän loukkaamisesta sekä lakia yksityisyyden suojasta työelämässä (2004/759). Viimeaikoina erityisesti on puhuttanut sähköisen viestinnän tietosuojalaki (2004/516) ja sen Lex Nokiaksi ristitty uudistus (HE 48/2008).
Maksuperustelaki	Valtion maksuperustelaki (1992/150)
Kansainväliset suositukset	OECD Recommendation of the Council concerning Access to Research Data from Public Funding (OECD 2006) OECD Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information (OECD 2008)
Tekijänoikeuslaki	Tekijänoikeuslaki (1961/404)
Kilpailulainsäädäntö	Kilpailulainsäädännön lähtökohtana on taloudellisen kilpailun ensisijaisuus verrattuna markkinoiden säätelyyn. Nykyisin voimassaoleva laki on vuonna 1992 säädetty laki kilpailunrajoituksista (1992/480).
EU Direktiivit	INSPIRE direktiivi (2007/2/EY) tähtää paikkatietojen käytön tehostamiseen, viranomaisten yhteistyön lisäämiseen ja monipuolisten kansalaispalvelujen syntymiseen. Suomessa direktiivin toteuttamisesta on säädetty laissa (421/2009) ja asetuksessa (725/2009) paikkatietoinfrastruktuurista. PSI (Public Sector Information) direktiivi (2003/98/EY) julkisen sektorin hallussa olevien tietojen uudelleenkäytöstä tähtää julkisen datan kaupallisen hyödyntämisen lisäämiseen.

3.1 Datan julkisuuteen vaikuttava lainsäädäntö

Hallinnon datavarantojen julkisuutta määrittää ensisijaisesti tietosuojalainsäädäntö ja viranomaisten asiakirjojen julkisuudesta määräävä ns. julkisuuslainsäädäntö. Kaikki tieto ei luonnollisesti ole eikä sen pidäkään olla julkista. Yksilöiden tietosuojan kunnioittaminen on keskeinen perusta, jolla voidaan säilyttää kansalaisten luottamus viranomaisiin kohtaan ja pienentää riskiä väärinkäytöksiin. Myös turvallisuus- tai muista syistä arkaluontoiseksi luokiteltu data ei ole julkista.

Julkisuuslain mukaan jokaisella on oikeus saada tieto viranomaisen julkisesta asiakirjasta ja muusta tallenteesta. Viranomaisen hallussa olevat asiakirjat ovat julkisia, jolle niiden julkisuutta ole välttämättömistä syistä laissa erikseen rajoitettu. Julkisen asiakirjan antamisesta ei peritä maksua muun muassa silloin, kun sähköisesti talletettu asiakirja lähetetään tiedon pyytäjälle sähköpostitse tai kun pyydetyn asiakirjan antaminen kuuluu viranomaisen neuvonta-, kuulemis- tai tiedotusvelvoitteen piiriin. Jos maksu peritään, kustannuksiin ei saa sisällyttää asiakirjan esille hakemisesta ja salassa pidettävien osien poistamisesta aiheutuvia kustannuksia, kun kysymyksessä on tavanomainen asiakirjapyyntö.

Tietosuojassa ei ole kysymys tiedon suojelemisesta, vaan kansalaisten oikeudesta elää ilman epäilyksiä heidän yksityiselämänsä koskevien tietojen julkittulemisen seuraamuksista. Laissa suojataan yksityisyyttä (kuten tietoja henkilön taloudellisesta asemasta, terveydestä ja poliittisesta kannasta) ja muita oikeuksia, kuten oikeutta vapaaseen liikkumiseen ja kokoontumiseen. Esimerkiksi matkakorttijärjestelmään tallentuivat aikanaan myös matkakortin leimaukset eri bussilinoilla. Niistä oli jopa mahdollista saada jälkikäteen tuloste itselleen. Tämä herätti keskustelua tietosuojasta nimenomaan liikkumisvapauden näkökulmasta ja leimausten tallentamisesta luovuttiin.

Tietosuojalainsäädännön kulmakivenä on henkilötiedon määrittely ja tarkoituksena on henkilötietojen suojaaminen oikeudettomalta ja henkilöä vahingoittavalta käytöltä. Henkilötiedon määrittely yksiselitteisesti on haastavaa. EU on määritellyt sen 26-sivuisessa dokumentissa (EU 2007). Henkilötiedon käsittää helpommin ajattelemalla sen vastakohtana anonyymia tietoa, joka ei ole yhdistettävissä henkilöön. Datavarantojen avaamisessa on helpointa lähteä liikkeelle anonyymien tietojen julkaisusta.

Henkilötietolaissa on yksityiskohtaiset säännökset henkilötietojen rekisteröimisestä ja käytöstä. Lähtökohtana on, että tietoja, joiden perusteella yksityinen henkilö voidaan tunnistaa, ei saa julkaista ilman kyseisen henkilön suostumusta. Tämä koskee myös välillisen tunnistamisen mahdollistavia tietoja. Joissain tapauksissa, esimerkiksi tutkimuskäyttöä varten voidaan luotetuille tahoille luovuttaa erillissopimuksilla myös sellaisia datavarantoja, joita ei voitaisi julkaista täysin vapaasti kaikkien saataville. Tällaisissa erillistapauksissa on ehdottoman tärkeää tehdä sopimuslisensointi läpinäkyvästi, jotta kuka tahansa voi perehtyä perusteisiin, joilla sopimus on syntynyt.

Aineistojen julkisuuteen liittyvä lainsäädäntö on melko yksiselittäistä ja helposti tulkittavaa, mutta joiltain osin aineistojen lainmukaista julkisuutta saatatta olla syytä rajoittaa, mikäli niiden saavutettavuus esimerkiksi verkkojakelun myötä dramaattisesti helpottuu. Nykyisin on mahdollista kaivaa esille paljonkin henkilöihin liittyvää julkista dataa, mutta se vaatii kohtuullista vaivannäköä ja aikaa. Mikäli kaikki sama data olisi saatavilla internetissä yhden napin painalluksella, saattaisi yksityisyydensuoja vaarantua.

Esimerkiksi Maanmittauslaitoksella on ollut tilanne, jossa samassa palvelussa on tietoa kiinteistöistä, niiden omistajista, tonteista ja metsästä. Tulkinta oli, ettei kiinteistön omistajan nimeä voi kertoa samassa palvelussa, vaikka kiinteis-

tön omistajan nimi onkin julkinen tieto. Syynä tähän oli, että palvelun kautta olisi liian helppo arvioida kiinteistön omistajan metsäomaisuuden rahallinen arvo, jonka katsotaan kuuluvan yksityisyyden suojan piiriin. On hyvä, jos tämänkaltaisiin asioihin varaudutaan ennakkoivasti niin että mukana uusien käyttömahdollisuuksien suunnittelussa ovat myös asioista vastaavat viranomaiset.

Tietosuoja on osa tietoturvaa, joka tarkoittaa yleisesti järjestelmän, datan, palveluiden ja tietoliikenteen suojaamista niiltäkin osin, missä tietosuojan kannalta keskeiset ihmisten perusoikeudet eivät ole uhattuina. Tietosuojavaltuutettu on esittänyt nykyisten hajallaan olevien tietoturvavelvoitteiden keskittämistä yleiseksi tietoturvalaiksi (YLE 2009). Yleisen ja teknologianeutraalin tietoturvalain tarve nousee keskeiseksi pilvipalveluiden - eli internetin kautta tarjottavien palveluiden - yleistyessä. Kun yritykset ja yksittäiset ihmiset alkavat laajemmassa määrin tallentaa omia tiedostojaan internetiin ja käyttämään pilvipalveluita, hankkivat he palvelut mieluiten luotettavina pidetyissä maissa toimivilta yrityksiltä. Mikäli tietosuojavaltuutetun ehdotus saa kannatusta, voisi Suomi toimia vastaavan lainsäädännön toteuttamiseksi myös EU:n tasolla. Tietoturvaa ei tämän oppaan puitteissa kuitenkaan tarkastella laajemmin.

3.2 Datan maksullisuuteen vaikuttava lainsäädäntö ja sääntely

Valtion maksuperustelain mukaan julkishallinnon suoritteet ovat maksullisia, ellei maksuttomuudelle ole perusteltua syytä. Maksuperustelaki antaa yleiset perusteet viranomaisten suoritteiden, kuten tietoluovutusten, palvelujen ja tavaroiden maksullisuudelle. Yksityiskohtaiset säännökset maksuista ja niiden suuruudesta määräytyvät asetuksilla eri ministeriöissä. Maksuperustelain mukaan julkisoikeudellisesta suoritteesta valtiolle perittävän maksun suuruuden tulee vastata suoritteen tuottamisesta valtiolle aiheutuvien kokonaiskustannusten määrää (omakustannusarvo). Laki kuitenkin mahdollistaa maksun perimättä jättämisen tai sen sen perimisen omakustannusarvoa alempana tai korkeampana.

Tietoluovutusten osalta voimassaolevaan maksuperustelakiin kohdistuu muutospaineita. (esim. VM 2007). Eurooppalainen PSI-direktiivi (luku 4.5.1) ja OECD:n suositukset, joiden tarkoituksena on, että maksu enimmillään koskisi vain irrottamisesta aiheutuvia marginaalikustannuksia. Valtiovarainministeriön antaman selvityksen (VM 2005) mukaan Suomi täyttää PSI-direktiivin vaatimukset lainsäädännön osalta. Direktiivin tavoitteita tukevia toimenpiteitä (julkisten tietojen saatavuuden läpinäkyvyyttä ja helpottamista mahdollisimman alhaisin kustannuksin, mieluummin ilmaiseksi) ei ole Suomessa toteutettu. OECD:n suosituksissa painotetaan julkisin varoin tuotetun tiedon kohdalla tekijänoikeuksien käyttöä tavalla, joka helpottaa uudelleenhyödyntämistä. Suositusten mukaan on esimerkiksi kannustettava tekijänoikeudellisista vaateista luopumista (Waiving of Copyrights). Lisäksi hinnoittelun pitäisi helpottaa saatavuutta ja uudelleen käyttöä eikä sen pitäisi ylittää marginaalikustannuksia.

3.3 Datan käyttöoikeuksiin vaikuttava lainsäädäntö

Tekijänoikeus ei suojaa ideoita eikä faktatietoja, vaan sitä muotoa, mikä niille annetaan. Tämä tarkoittaa, että faktoista koostuvaa tietosisältöä itsessään ei

omista kukaan, ellei sopimuksessa tiedon tuottajan kanssa ole niin määritelty. EU:ssa tekijänoikeuslaissa suojataan myös tietokantoja joko kirjallisina teoksina tai niin sanotun sui generis -suojan avulla. Mikäli tietokanta on suojattu kirjallisena teoksena suojan kohteena on tietojen omaperäinen valinta ja järjestäminen tietokantaan. Sui generis -suoja taas kohdistuu tietokannan keräämiseen, valmistamiseen tai esittämiseen käytettyyn huomattavaan työmäärään. Kirjallisten teosten tekijänoikeussuoja on 70 vuotta tekijän kuolinvuodesta ja sui generis -suoja-aika on 15 vuotta tietokannan valmistumisesta tai viimeisestä muutoksesta.

Mikäli aineisto ei kuulu tekijänoikeuksien piiriin, on erillisellä tiedon tuottajan ja käyttäjän välisellä luovutussopimuksella mahdollista kontrolloida aineiston uudelleenkäyttöä ja uudelleenjakelua. Käänteisesti, vaikka aineistoon kohdistuu tekijänoikeus, tämä ei tarkoita sitä, että sen käyttöä olisi valvottava tai lisensoitava. Tekijä voi luopua kaikista aineistoon automaattisesti kohdostuvista oikeuksista ilmoittamalla sen suoraan aineiston jakelun yhteydessä (Waiver of Rights). Aineisto voidaan asettaa vapaasti verkkoon saataville esimerkiksi Creative Commons nolla -lisenssillä (Liite 4).

Case: Vaaliraharaportti

Vaalirahakuuhunnan keskellä Lauri Tarastin johtama vaali- ja puoluerahoitus-toimikunta valmisteli ehdotuksen uuden vaalirahoituslain pohjaksi (OM 2009). Ehdotuksesta uutisoitiin laajasti ja mm. rahoituksen ylärajoihin liittyen käytiin keskustelua verkossa ja lehdistössä. Itse ehdotusteksti löytyy julkisesti oikeusministeriön sivuilta tavanomaisena PDF-dokumenttina. TKK:n tutkimusryhmä olisi halunnut uudelleenjulkaista alkuperäisraportin erillisessä www-palvelussa, jossa raporttia olisi voinut kappaleen tasolla kommentoida ja yksittäisiin kappaleisiin olisi voinut linkittää. Tutkimusryhmän pysäytti kysymys tekijänoikeuksista: saako näin tehdä? Liittyykö ministeriö raportteihin tekijänoikeuksia, jotka estävät julkaisun?

Kummassakin tapauksessa aineiston tuottajalla on tekijänoikeuksien suhteen kaikki avaimet käsissään kannustaa aineiston uudelleenkäyttöön. On tärkeää tuntea hallinnon tuottaman informaation nykytilanne suhteessa tekijänoikeuksiin, koska tekijänoikeudet syntyvät automaattisesti ja ne saattavat rajoittaa aineiston käyttöä enemmän kuin tekijä toivoisi. Erilaiset avoimet lisenssit (esim. Creative Commons, Science Commons ja Open Database Licence) asettavat teoksen käytölle mahdollisimman vähän rajoituksia.

Tietoaineistojen uudelleenjakamisesta on julkishallinnon eri toimijoiden välillä kirjavia sopimuskäytäntöjä, joissa saatetaan sanoa esimerkiksi, että «aineistoa saa jakaa, mutta ei massiivisissa määrin». Sopimukset ovat usein vanhoja eikä niitä tehtäessä ole huomioitu internetiä digitaalisen tiedon jakelukanavana, vaan tiedot on sopimuksella luovutettu esimerkiksi jonkin julkaisun tekoa varten. Monilta epäselvyyksilyä välttää, jos kuntien ja valtion organisaatioissa selkeästi määritellään suhtautuminen immateriaalioikeuksiin oman toiminnan tavoitteista käsin.

3.4 Kilpailulainsäädäntö

Kilpailunrajoituslain tarkoituksena on turvata terve ja toimiva taloudellinen kilpailu vahingollisilta kilpailunrajoituksilta. Kilpailunrajoituksia ja niiden vaikutuksia selvittävät etupäässä kilpailuvirasto ja lääninhallitus. Lain säännökset koskevat myös valtion virastoja ja laitoksia, jotka toimivat hyödykemarkkinoilla. Valtiovarainministeriön muistiossa (11/2004) kannustetaan hallinnon organisaatioita kuitenkin pyrkimään siihen, että ne eivät toimisi hyödykemarkkinoilla: "Jos

viranomaisen tuottamat tiedot ovat saatavissa markkinoilla, valtion tulee harkita oman tiedonkeruunsa tarpeellisuutta”.

Julkishallinnon organisaatioiden osallistuminen tietomarkkinoille saattaa vääristää kilpailua. Kilpailuhäiriöitä syntyy, jos kyse on aidosta markkinasuorituksesta eli samat tiedot tai palvelut on saatavissa myös muista lähteistä ja muihin tarkoituksiin. Jos viranomainen tuottaa liikeloudellisesti hinnoiteltavia suoritteita määräävässä markkina-asemassa, on suoritteiden hinnoittelussa otettava huomioon kilpailunrajoituksista annetun lain (480/1992) määräävän markkina-aseman väärinkäyttöä koskevat säännökset.

Julkisen datan jalostusketjussa on noudatettava kilpailulainsäädännön periaatteita. Kilpailulainsäädännön noudattaminen merkitsee sitä, että markkinoilla toimivien yksityisten yritysten on saatava ”raakadata” tai ”puolivalmisteet” samalla hinnalla kuin millä niitä siirretään viranomaisten välillä. Tämä asettaa suuret vaatimukset hinnanmuodostuksen avoimuudelle ja virastojen laskenta-toimelle. Erityisesti tämä edellyttää liikeloudellisesti hinnoiteltujen palvelujen erillistä seuranta kirjanpidossa ja kehittyntä kustannusseuranta. (VM 2003)

3.5 Julkishallinnon dataan liittyvät EU-direktiivit

Julkishallinnon avoimeen dataan liittyy merkittävästi kaksi EU:n direktiiviä: vuonna 2003 voimaan tullut julkisen datan kaupalliseen hyödyntämiseen tähtäävä PSI-direktiivi sekä vuonna 2007 säädetty paikkatietojen yhteiskäyttöä koskeva INSPIRE-direktiivi. Näiden kahden direktiivin tavoitteet ja toteutuskeinot sisältävät olennaisilta osilta sen, mitä tässäkin oppaassa käsitellään. PSI-direktiivillä on toistaiseksi ollut Suomessa vähäinen vaikutus ja INSPIRE-direktiivin toimeenpääntö on vasta käynnistynyt. Aika näyttää, mitä parannuksia se tuo tullessaan.

3.5.1 Julkisen datan kaupallinen hyödyntäminen ja PSI-direktiivi

PSI-direktiivi on Suomessa varsin heikosti tunnettu verrattuna INSPIRE-paikkatietodirektiiviin. Komission arvion mukaan direktiivi on koko unionin alueella merkittävästi edistänyt julkistiedon EU-sisämarkkinoiden muodostumista, mutta toimeenpanon edistymistä täytyy edelleen seurata tarkasti (EY Komissio 2009). Mikäli vuonna 2003 säädetyn PSI-direktiivin tavoitteita olisi Suomessa kunnolla edistetty, ei tätä opasta enää tarvittaisi. Datavarantojen kaupallisen hyödyntämisen edistäminen tukisi myös epäkaupallista käyttöä ja viranomaisten välisen tiedonvaihdon kehittymistä. Direktiivi sisältää artikloja, joiden toteuttaminen edistäisi merkittävästi mm. lisensointikäytäntöjen selkeytymistä ja julkisen tiedon löydettävyyttä.

Käyttöehdoista (artikla 7) direktiivissä sanotaan: ”Kaikki julkisen sektorin elinten hallussa olevien asiakirjojen uudelleenkäyttöä koskevat ehdot ja vakiomaksut on vahvistettava etukäteen ja julkistettava, mahdollisuuksien mukaan ja tarkoituksenmukaisissa tapauksissa sähköisessä muodossa. Julkisen sektorin elimen on pyynnöstä ilmoitettava julkistetun maksun laskentaperusteet.”

Julkisen tiedon löydettävyyttä koskien (artikla 9) direktiivissä sanotaan: ”Jäsenvaltioiden on varmistettava, että on toteutettu uudelleenkäytettävissä olevien asiakirjojen hakua helpottavia käytännön järjestelyjä, kuten mahdollisuuksien mukaan Internetin välityksellä saatavilla olevia luetteloja tärkeimpien asiakirjojen sisältövaroista ja

hajautettuihin sisältövaraluetteloihin linkitettyjä portaalisivustoja.”

Tietoluovutusten hinnoittelussa direktiivi määrittelee perittävien maksujen katon ja kannustaa soveltamaan tietojen irrottamisen ja luovuttamisen teknisiin kustannuksiin perustuvaan hinnoitteluun. Digitaalisen tiedonsiirron ja kopioimisen tapauksessa julkaisun laskuttamisesta saattaa syntyä suuremmat kustannukset, kuin itse irtiotosta. Hinnoittelumääräys oli jo direktiivin julkaisun aikaan kompromissi ja sen jälkeen argumentaatio julkisen datan maksuttoman tarjoamisen puolesta on Euroopassa voimistunut (ks luku 4).

Valtvarainministeriön perustietovarannot-projektin selvitysten (VM 2007) mukaan PSI-direktiivin toimeenpanemiseksi Suomessa on muutettu julkisuuslain pykälää 34 (HE 20/2005) niin, että virastot veloitetaan määrittelemään etukäteen tietoluovutuksista perittävät maksut ja julkaisemaan ne. Muita direktiivin tavoitteita tukevia toimenpiteitä, kuten esimerkiksi artiklan 9 mukaisia datakatalogeja ei Suomessa ole tähän mennessä toteutettu. Viranomaiset ovat toteuttaneen sovellus-sovellus-yhteyksiä joillekin lisäarvopalvelujen tuottajille. Näitä yhteyksiä ei kuitenkaan aktiivisesti tarjota suoraan muille toimijoille. Yksittäisten viranomaisten verkkosivuilta ei myöskään usein helposti selviä, mitä datavarantoja (kaupalliseen) hyödyntämiseen olisi saatavissa.

3.5.2 INSPIRE-direktiivi ja paikkatietolaki

Paikkatietoaineistot käsittävät merkittävän osan julkishallinnon datavarannoista ja niiden avaamisessa on edistytty pisimmälle, aina lainsäädännön tasolle saakka. INSPIRE-paikkatietodirektiivin myötä ympäristöä kuvaavat paikkatietoaineistot tulevat olemaan saatavilla yhtenäisessä muodossa kaikista EU:n jäsenmaista. Paikkatietolailla viitataan 17.6.2009 voimaan tulleeseen lakiin (421/2009) ja asetukseen, jotka toteuttavat paikkatietodirektiivin Suomessa. Laki pitäytyy Inspire-direktiivin vaatimuksissa eikä laajenna sen soveltamisalaa.

Aineistolla tarkoitetaan sähköisessä muodossa olevaa tunnistettavaa tietojen kokonaisuutta, joka voidaan julkaista. Paikkatietolakia tukevassa asetuksessa on määriteltä, mitä paikkatietoaineistoja laki koskee. Mukana on parikymmentä valtionhallinnon virastoa, laitosta, eräitä aluehallinnon organisaatioita ja kunnat. Kokonaisuuteen kuuluu runsaat sata valtakunnallista tietoaineistoa ja lisäksi lukuisia kuntien aineistoja. Yhteensä lain tarkoittamia aineistoja on Suomessa noin pari tuhatta.

Lain mukaan paikkatietoaineistoja hallinnoivien viranomaisten tulee kuvailla aineistot metadatan avulla ja liittää ne hakupalveluun. Tämän lisäksi yhteiskäyttöön tarkoitetut versiot aineistoista pitää asettaa katseltavaksi ja ladattavaksi internetiin. Aineistojen maksullisuuteen liittyvissä kysymyksissä laki viittaa valtion maksuperustelakiin, mikä tarkoittaa, että maksuperustelakia mahdollisesti uudistettaessa muutokset koskevat myös paikkatietoaineistoja. Kuitenkin metatiedon käyttö on kaikille maksutonta ja katselupalvelujen mahdollinen maksullisuus tulee erikseen perustella. Mikäli maksuja peritään, on toteutettava sähköinen asiointi ja maksaminen. Käyttöä koskevien ehtojen ja sopimusmallin tulee olla verkossa saatavilla.

Paikkatietolainsäädäntö ja sen toimeenpanon käytännöt soveltuisivat monilta osin suoraan malliksi myös muiden julkishallinnon datavarantojen avaamiseen. Tuoreen lain toimeenpano on vielä oppaan kirjoitushetkellä kesken. Ensimmäisen vaiheen paikkatietoaineistojen metatiedot pitää olla toimitettu kansalliseen hakupalveluun (www.paikkatietohakemisto.fi) 1.12.2010 mennessä. Paikkatietojen katselupalvelujen tulee olla saatavilla keväällä 2011 ja latauspalveluiden vuoden 2011 lopulla. Lain toimeenpanoa koordinoi Maanmittauslaitos. Direktiivin toimeenpanon seuraaminen on tämän oppaan teeman kannalta erittäin mielenkiintoista.

*Maksuperustaisuus johtaa valitettavasti
erittäin laadukkaiden datavarantojen
vajaakäyttöön ja jopa saman tiedon
keräämiseen monta kertaa eri paikkoihin.*



4. Avoimen datan taloudellisia näkökohtia

Tämä luku kokoaa yhteen keskusteluja, väitteitä ja näkökulmia, joita on esitetty julkisen datan hinnoittelusta. Tähän teemaan liittyviä asioita ovat Suomen voimassa oleva maksuperustelaki, EU:n suosittama korkeintaan jakelusta aiheutuneisiin kustannuksiin perustuva hinnoittelu, hallinnon edistämä kokonaisedullisuusajattelu sekä kansainväliset ajattelutavat avoimesta datasta kansallisen innovaatiostrategian yhtenä mahdollistajana. Lopuksi tarkastelemme avoimen datan mahdollisia rahoitusmalleja. Avoimen datan kokonaishyötyjä ja kumulatiivisia vaikutuksia on laskettu vielä kansainvälisestikin vähän ja taloutta käsitellään tässä pääosin esimerkkien kautta.

Julkishallinnon datan maksuttomuuteen liittyvissä keskusteluissa on toistuvasti noussut esiin kaksi ajatusta: toistaalta avoimen datan oletetusti positiiviset taloudelliset vaikutukset ja toistaalta huoli aineistotuotannon rahoittamisesta.

Avoimen datan taloudelliset hyödyt perustuvat ajatukseen datasta julkishyödykkeenä. Aiemmin tiedon jakelua rajoittivat fyysisestä kopioinnista, paperille painamisesta ja postituksesta tai muusta logistiikasta aiheutuvat jokaista uutta kopiota koskevat marginaalikustannukset. Verkossa jaettavan digitaalisen tiedon tapauksessa yhdestä yhden uuden kopion aiheuttamat marginaalikustannukset syntyvät lähinnä bittien siirrosta tietoverkoissa ja ovat häviävän pieniä. Eikä jaettu data häviä tai kulu, vaikka sitä kuinka paljon kopioitaisiin.

Ideoinnin, keksintöjen ja innovaatioiden pohjalla on aina informaatiota, joten nämä toiminnot liittyvät hyvin suoraan avoimesta datasta käytävään keskusteluun.

Yhdysvalloissa on vahvat sananvapauden ja valtion vähäisen markkinoille osallistumisen perinteet. Siellä on katsottu, että verovaroin tuotettu data on syytä tarjota kaikkien saataville – internetin välityksellä myös USA:n ulkopuolelle. Eurooppalaista maksuperustaisuus-mallia sovelletaan nykyisin muun muassa monien paikkatietojen jakeluun. Siinä datavarannon tuottaja ja ylläpitäjä kattavat suuren osan kuluistaan datan jakelusta saamallaan tuloilla. Suomessa käytiin 1980-luvulla laaja keskustelu informaation hinnoittelusta. Se tuotti 1990-luvun alussa nykyisen maksuperustelain. Aiheeseen palattiin 1990-luvun puolivälissä, kun Suomi alkoi luoda tietoyhteiskuntastrategioita, mutta perusnäkemys säilyi (Kuronen 1998a).

Maksuperustaisuus johtaa valitettavasti erittäin laadukkaiden datavarantojen vajaakäyttöön ja jopa saman tiedon keräämiseen monta kertaa eri paikkoihin (Benson 2009). Toisaalta hinnoittelu rajoittaa dataan perustuvien tuotteiden kehittymistä. Esimerkiksi säädätan uudelleenkäyttöön perustuvien riskinhallintatuotteiden markkina oli vuonna 2002 Yhdysvalloissa 50 kertaa suurempi kuin Euroopassa (Weiss 2002).

Maksuperustaisuuden vaihtoehtona datan jakelun yhteydessä puhutaan maksuttomuudesta tai enintään irtiottokustannuksiin perustuvista maksuista. Maksuttomuus tarkoittaa, että datavarantojen tuotto, ylläpito ja jakelu hoidetaan kokonaan budjettivaroin. Irtiottokustannuksiin (VM 2003) perustuva hinnoittelu puolestaan tarkoittaa, että datan kerääminen ja ylläpito kustannetaan budjettivaroista, mutta datan käyttäjät maksavat datan jakelusta aiheutuneet lisäkustannukset.

Datan maksuton saatavuus kaikille on tärkeä keino lisätä sen avoimuutta. Maksuttomuus ei tarkoita, etteikö avoimen datan ympärille voisi kehittyä liiketoimintaa. Hyvä esimerkki tästä on avoimen lähdekoodin projektit. Ohjelmat ovat käyttäjilleen maksuttomia, mutta silti niiden ympärille on kehittynyt liiketoimintaa muun muassa tarjoamaan ohjelmistojen asennusta, käyttäjätukea,

ylläpitoa sekä räätälöityä versiointia.

Avoimen datan ajattelussa Suomi hyötyy enemmän maksuttomuudesta kuin nykyisen maksuperustelain mukaisesta suurelta osin hallinnon sisäisestä aineistokaupasta tai aineiston tuotantokustannuksiin nähden pienien irtiotto-kustannusten veloittamisesta. Pientenkin maksujen keräämiseen vaaditaan byrokratiaa, jonka ylläpitäminen maksaa. Liiketaloudellisen näkökulman lisäksi maksuttomuus on demokratiaa. Ihmiset saavat tasavertaisesti ja maksutta argumenttiansa tueksi tarvitsemaansa dataa, jota he voivat jalostaa ja käyttää parhaaksi katsomallaan tavalla.

Case: Hallinnon sisäiset tietoluovutukset

«Vilkas kaupankäynti julkisilla tiedoilla on siirrellyt valtion rahaa taskusta toiseen, mutta nettotulot eivät ole lisääntyneet» (Kuronen 1998a)

Huomattava osa tietojen siirrosta viranomaisten välillä on itseasiassa vastikkeetonta. Tämä perustuu yleensä siihen, että tietoja vastaanottavaa viranomaista koskevassa laissa määrätään, että se saa toisen viranomaisen tietoja maksutta käyttöönsä. Tällaisia määräyksiä on mm. kansaneläkelaisissa, poliisilaissa, tilastolaissa ja tullilaissa. Tästä huolimatta lähes puolet viranomaisten tietoluovutusten tuloista tulee toiselta viranomaiselta. Valtiovarainministeriö on tehnyt perusteelliset selvitykset hallinnon sisäisten tietoluovutusten hinnoittelusta vuosina 2003 ja 2004 (VM 2003, VM 2004). Jälkimmäisessä muistiossa on myös viisi toimenpideehdotusta, joita ei kuitenkaan ole toteutettu. Ensimmäinen ehdotus on siirtyminen enintään irroituskustannuksiin perustuvaan hinnoitteluun.

Selvityksessä tehtiin kysely 17 suurimmalle digitaalisen tiedon tuottaja- ja jakajaorganisaatiolle valtion hallinnossa. Tietoluovutuksista saadut tulot kyselyssä mukana olleilla viranomaisilla vuonna 2002 olivat yhteensä lähes 28,2 miljoonaa euroa. Tietoluovutusten menoiksi vastaajat ilmoittivat yli 13,5 miljoonaa euroa, jolloin nettotuloksi jäi yhteensä 14,6 miljoonaa euroa. Tietojen luovuttamisesta saatujen tulojen merkitys vaihtelee toimijoiden välillä (Taulukko 4.1).

Taulukko 4.1 Digitaalisten tietoluovutusten tulot eräissä viranomaisissa vuonna 2002 (VM 2004)

	Valtio	Kunta	Yritykset ja kansalaiset	Yhteensä (1000 €)	Osuus maksutuloista
Väestörekisterikeskus	2528	1223	4771	8522	98,0%
Ilmatieteen laitos	4454	66	2127	6647	86,3%
Ajoneuvo-hallintokeskus	330	229	5017	5576	10,0%
Maanmittauslaitos	2728	236	1332	4296	9,0%
Verohallinto	757	151	87	995	27,0%

Vuonna 2007 aiempia selvityksiä tarkasteltiin uudelleen ja todettiin: "Vuoden 2004 jälkeen muutoksia luovutushinnoitteluun on tehty liikenne- ja viestintäministeriön hallinnonalalla. Digiroad-tietojärjestelmästä tietoja luovutetaan irrotuskustannuksella ja samoin toimitaan ajoneuvoliikennejärjestelmän osalta. Muuten maksuperusteet ovat pysyneet ennallaan" (VM 2007).

4.1 Muuttunut toimintaympäristö

Digitaalinen data, informaatio ja tieto eroavat muista tuotannontekijöistä. Niitä voidaan muokata, kopioida ja jakaa ja silti alkuperäinen lähde säilyy sen omistajalla ja tuottajalla (Hintikka 1993). Laadukkaiden datavarantojen kerääminen, ylläpito ja jakeluinfrastruktuurin rakentaminen ei luonnollisesti ole ilmaista, vaikka jakeluun ja kopiointiin liittyvät marginaalikustannukset ovatkin digitaalisuuden myötä pienentyneet. Kustannusrakenne on painottunut jakelumediasta tuotantoon ja järjestelmäinvestointeihin.

Internetin yleistymisen myötä aletaan vasta havahtua siihen, mitä tämä käytännössä tarkoittaa. Informaation kaupalliseen tuottamiseen ja myymiseen erikoistuneet toimialat, kuten tietosanakirjat sekä joukkoviestimet ovat historian- ja pahimmassa taloudellisessa kriisissä. Samaan aikaan internetissä maksuton informaatio, tuotteet ja palvelut muodostavat erittäin kannattavan bisneksen, kuten Googlen liiketoiminta osoittaa.

Wired-lehden päätoimittaja Chris Anderson (2008) käyttää tästä murroksesta termiä ilmaistalous (freeconomics). Laajakaista sekä internet-palveluiden talennuskapasiteetin kasvu ovat yhdessä johtaneet siihen, että monissa internet-palveluissa on kannattavampaa tarjota yksittäisen käyttäjän maksuton peruskäyttö verrattuna laskutuksesta ja transaktioista koituviin kuluihin. Maksuton peruskäyttö rahoitetaan muilla ansaintamalleilla, joita Andersonin mukaan ovat muun muassa mainosrahoitteisuus sekä korvaavuus (subsidiary), eli toiminta rahoitetaan esimerkiksi konsultoinnilla tai muilla kuukausimaksullisilla lisäpalveluilla.

Kulurakenteen muutokset koskevat myös digitaalisia palveluita. On mahdollista esimerkiksi muodostaa joukkoliikenteen aikataulutietokannan perusteella käyttäjille reittiehdotuksia ilman ihmistyövoimaa. Työvoimakustannukset siirtyvät ylläpitoon ja palvelun kehittämiseen. Tästä seuraa, että palvelun kapasiteettia voidaan skaalata työvoimakustannuksista riippumatta ja palvelua voidaan tarjota kellon ympäri.

Digitaalinen tiedonvälitys on arkipäiväistynyt hämmästyttävän nopeasti ja muuttanut merkittävällä tavalla ihmisten arkea, eikä muutosvauhti ainakaan näytä hidastuvan. Vielä 90-luvun alkupuolella käytiin julkishallinnossakin keskusteluita siitä, pitäisikö kaikilla virastoilla olla omat verkkosivut. Vastaavalla tavalla nyt keskustellaan virkamiesten läsnäolosta sosiaalisessa medissa ja koneluettavan datan jakamisesta maksutta kaikille. Digitalisoitumisen mukanaan tuomat muutokset ovat niin suuria, että ne muuttavat prosesseja laadullisesti. Nykyisin ei tehdä vain samaa vanhaa enemmän ja nopeammin, vaan toimijoiden roolit muuttuvat ja ihmiset tekevät automaation myötä kokonaan eri asioita.

Verottaja on yksi edistyksellisimmistä toimijoista Suomessa. Se on osannut ottaa hyödyn irti digitalisoitumisesta muuttamalla sen myötä myös omia toimintamallejaan. Siinä, missä aiemmin kaikki kansalaiset tuskailivat veroilmoituksen teon kanssa tammikuussa, uudisti verottaja koko prosessin siten, että suurin osa kansalaisista saa verottajalta valmiin veroehdotuksen, johon tehdään vain tarvittavat muutokset. Tämä säästää sekä verottajan resursseja että kaikkien veronmaksajien aikaa ja hermoja. Digitaalisuuteen olisi voitu siirtyä myös muuttamatta prosessia esimerkiksi vain mahdollistamalla perinteisen veroilmoituksen täyttäminen nettilomakkeella, mutta kokonaishyöty olisi ollut tällöin pienempi.

4.2 Maksuperustelaki ajalta ennen internetiä

Suomen nykytilanne hallinnon tuottaman datan maksullisuudesta perustuu pitkälti maksuperustelakiin (1992/150), joka luotiin vuonna 1992, ennen internetin mukanaan tuomaa informaation jakelurakenteen muutosta. Kun informaatiosyhteiskunnan ensimmäinen aalto saapui Suomeen 1980-luvulla, hallinnon tuottamasta datasta käytiin vilkasta keskustelua. Toisaalta tuolloin viitattiin Yhdysvaltain malliin, jossa verovaroin tuotettua dataa pidetään yleishyödykkeenä. Toisaalta ajateltiin, että tuotetulla datalla pitäisi olla hinta, joka muodostuu sen tuottamisen kustannuksista. Lisäksi hallinnon tuottamalle datalle haluttiin ainakin minimihinta myös siksi, etteivät kansalaiset vaivaisi turhaan hallintoa pyynnöillään. Hinnan, olipa se pieni tai suuri, katsottiin vähentävän pyyntöjä ja työajan käyttöä pyyntöjen täyttämiseen.

Valtion maksuperustelain säätämisen taustalla oli muun muassa 1990-luvun alun lama ja siihen liittyvät julkisen hallinnon muutospaineet. Viranomaisilta edellytettiin yhä kustannustehokkaampaan toimintaan ja tämän johdosta tuli myös tarve viranomaisten palvelujen hinnoittelulle. Maksuperustelaki ja netobudjetoinnin toteuttaminen yhdessä vaikuttivat voimakkaasti viranomaisten palvelujen tuottamis- ja jakelutapaan. 1990-luvulla lakia yritettiin muutamia kertoja muuttaa, mutta uudistus ei edennyt valtioneuvoston käsittelyyn, joten laki on ollut nykyisten käytäntöjen muovaajana jo pitkään. Valtion toimintojen liikelaitostaminen tuki tätä kehitystä (VM 2003).

Maksuperustelaki sääntelee viranomaisten maksullista toimintaa yleensäkin ja antaa yleiset perusteet eri suoritteiden, kuten palveluiden ja tavaroiden, mutta siinä sivussa myös tietoluovutusten maksullisuudelle (luku 4). Lain asettamisen aikaa myös tietoluovutusten maksullisuus oli sinänsä perusteltua. Dataa piti kaivaa esiin, tulostaa ja postittaa jne. Nykyään alkuperäisen maksuperustelain oletama toimintaympäristö on digitaalisen datan osalta tyystin muuttunut. Hallinto voisi yksinkertaisesti ja automatisoidusti laittaa datavarantonsa tarjolle internetiin ilman tulostelua ja postituksia. Tietoja luovuttavien organisaatioiden on toki investoitava tietopalvelujärjestelmiin, mutta itse kustannusperiaate ei enää ole perusteltu tältä osin.

Eräät tietojen luovuttajat arvioivat, että siirtyminen tietoluovutusten omakustannushintaan heikentäisi niiden mahdollisuuksia kehittää ja ottaa käyttöön uusia palveluja. Nykyinen tilanne ei tätä olettamaa kaikilta osin tue, vaan on nähtävissä, että korkeita maksuja perivien virastojen tietopalvelut eivät välttämättä ole monipuolisempia ja uudenaikaisempia kuin muidenkaan vastaavat palvelut (VM 2007).

4.3 Irtiottokustannus EU:n maksimisuosituksena

Lainsäädännön on vaikea pysyä mukana nykyisenkaltaisessa nopeassa muutoksessa. Vuonna 2003 säädetyssä PSI-direktiivissä (ks 3.5.1) on lähtökohtana, että maksuperustaisuudesta luovutaan ja datan luovuttamisesta peritään enintään irtiottokustannukset. Ne olivat direktiivin valmistelun aikana 2000-luvun alussa vielä jossain määrin perusteltuja, kun esimerkiksi cd-rom oli yleinen tapa julkaista aineistoja.

Direktiivissä ei tarkkaan ottaen edes määritellä irtiotto- tai irroituskustannuksia, vaan puhutaan jäljentämis- ja jakelukustannuksista, mikä vielä paremmin osoittaa ajattelun vanhakantaisuuden: «Tässä direktiivissä säädetty kulujen yläraja ei vaikuta jäsenvaltioiden tai julkisen sektorin elinten oikeuteen periä alhaisempia maksuja tai olla

kokonaan perimättä maksua, ja jäsenvaltioiden olisi rohkaistava julkisen sektorin elimiä asettamaan asiakirjoja saataville hintaan, joka ei ylitä asiakirjojen jäljentämisestä ja jakelusta aiheutuvia lisäkuluja.»

Direktiivin suurin motiivi on PSI-markkinoiden koon kasvattaminen ja etenkin yritysten saaminen uudelleenkäyttämään julkishallinnon tuottamaa informaatiota. Euromääräisesti laskettuna informaatiomarkkinoiden koko saattaa kasvaa, kun hintoja alennetaan ja siten mukaan saadaan uusia yrityskäyttäjiä. PSI-direktiivi vie asioita oikeaan suuntaan, mutta se on oman aikakautensa tuote. Digitaalisen maailman kehityksessä 7 vuotta on iäisyys, ja jos lasketaan direktiivin valmistelun alusta noin vuodesta 2000, sijoittuu se aikajanelle ensimmäisten nettisivujen ja nykypäivän puoliväliin. Direktiivissä ei vielä otettu huomioon laadullista muutosta, vaan siinä haluttiin lisätä, nopeuttaa ja helpottaa julkishallinnon tietovarantojen uudelleenkäyttöä sellaisen prosessin mukaisesti, millä se oli ennenkin tapahtunut. Pääasiassa tämä tarkoitti sitä, että yritykset sopimuksella saivat käyttöönsä aineistoja.

Suomen tie- ja katuverkkoa kuvaavan Digiroad-aineiston saa käyttöönsä nykyisin pelkällä irtiottokustannuksella, joka on joitain satoja euroja koko aineistolta. Tämä on dramaattinen ero verrattuna maksuperustaisiin rekistereihin. Esimerkiksi Ajoneuvorekisterin koneellisista suorakyselyistä veloitetaan 3,50 euroa hakua kohden, jolloin aineiston kokonaishinnaksi tulee miljoonia euroja.

Nykyisin Digiroad-aineisto toimitetaan tilaajalle vielä vanhanaikaisesti DVD-ROM-levyllä. Verkon pilvipalveluita hyödyntäen on tällaisen datamäärä siirtäminen tilaajalle nykyään mahdollista tehdä alle euron kustannuksilla ilman minikäänlaisia alkuinvestointeja tai ylläpitokustannuksia. Vaikka hinta ei pidäkään sisällään aineiston päivittämisen automatisoimista eikä mahdollisesti tarpeellista käyttäjien rekisteröitymistä. Herää kysymys, eikö asiaa voitaisi viedä loppuun, kun on päästy niin lähelle maksuttomuutta. On helppo kuvitella, että tuo viimeinen pieni maksu ja siihen liittyvä kirjallisten sopimusten teko ja byrokraatia helposti rajoittavat merkittävän osan etenkin kokeilevasta ja uutta luovasta aineiston käytöstä pois.

4.3.1 Hinnoittelun vaikutus datan uudelleenkäyttöön

Yleisesti ottaen datan hyödyntämiseen liittyvät ajalliset ja rahalliset kustannukset rajaavat sen hyödyntämismahdollisuuksia. Hinnoittelun vaikutukset saattavat erota kuitenkin merkittävästi toimijoiden välillä. Hallinnon organisaatiot ja vakiintuneet yritykset ovat tottuneita ostamaan ja neuvottelemaan datan hyödyntämisestä. Uuden innovatiivisen toiminnan kehittämisessä - kuten startup-yritysten toiminnassa, tutkimuslaitoksissa ja järjestö- ja kansalaissektorilla tilanne on kuitenkin toinen. Kun datan hyödyntämistä haluttaisiin kokeilla, saattavat kohtuullisetkin irtiottokustannukset ja niistä aiheutuva sopimusbyrokraatia vähentää kokeilunhalua, koska ei ole vielä varmuutta siitä, onko aineistolle pidemällä aikavälillä käyttöä.

Niiden yritysten näkökulmasta, joilla on jo toimivia julkiseen dataan pohjautuvia palveluita ja liiketoimintamalleja, ei ole suurtakaan merkitystä sillä, onko data saatavissa edullisesti irtiottokustannuksilla vai täysin maksutta. Tärkeintä on, että hinnoittelu- ja sopimuskäytännöt ovat selkeät ja datan ostaja saa sopimuksen turvin takuun datarajapinnan palvelutasosta. Se on usein yrityskäytössä niin kriittistä, että siitä halutaan maksaa. Rajoittuneemman kilpailun mielessä tällaisille vakiintuneille yrityksille saattaa olla jopa parempi, ettei data ole täysin maksutonta eikä saatavuus liian helppoa.

Tutkimus- ja koulutuskäyttö on konkreettinen esimerkki tilanteesta, jossa pelkkä aineistojen halventuminen ei vaikuttaisi niiden käyttöön, vaan vasta hintojen poistuminen kokonaan. Tutkimusorganisaatioissa on tyypillistä, että tut-

kijat haluaisivat joustavasti tehdä kokeiluja data-aineistojen hyödyntämisestä, mutta heillä ei ole käytännössä omaa valtuutusta tehdä pieniäkään hankintoja. Tämän takia maksu pitää perustella ja viedä raskaan byrokratian läpi. Kokeilut jäävät usein tekemättä, mikäli aineisto ei ole saatavissa maksutta.

Tutkimus- ja koulutuskäyttöä varten on kehitetty erilaisia ratkaisuja, joissa aineisto on saatavilla rajatulle käyttäjäryhmälle maksutta. Esimerkiksi Tieteen Tietotekniikan keskuksen (CSC) PaITuli-paikkatietopalvelu tarjoaa paikkatietoaineistoja suomalaisten korkeakoulujen tutkimus- ja opetuskäyttöön. Tämä kuitenkin vaatii aina rajanvetoa siitä, mikä on tutkimuskäyttöä. Kun tutkimuksen ja tutkimustulosten liiketoiminnallisen hyödyntämisen välistä raja-aitaa pyritään samanaikaisesti madaltamaan, on rajanveto turhauttavaa. Myöskin aloittavien yritysten mahdollisuuksia hyödyntää aineistoja pitäisi parantaa. Tätä varten on esitetty nk. start-off sopimuksia, jonka avulla aineistoja voitaisiin tarjota nuorille yrityksille ja korvaukset tulisivat ajankohtaisiksi vasta siinä vaiheessa, kun aineisto alkaisi tuottaa tuloja (Hermans 2009).

Pitkällä tähtäimellä yllämainittuja asiakasryhmiin tai käyttötarkoituksiin kohdistuvia hintadiskriminaatiomalleja ei kuitenkaan voida suositella niiden hyvistä tarkoituseristä huolimatta. Ne eivät tue kokonaisvaltaisesti joustavaa aineistojen hyödyntämistä ja useampien erilaisten sopimuskäytäntöjen ylläpito vain rasittaa sekä dataa tarjoavia että hyödyntäviä toimijoita. PSI-direktiivissä myös erityisesti pyritään eroon kaikista yksinoikeudellisista ja syrjivistä hinnoittelusta ja sopimuskäytännöistä.

Monissa tapauksissa ei selkeästi tiedetä, paljonko aineiston myynnistä tai irtiottoperustaisista veloituksista todella kertyy tuloja ja toisaalta paljonko aineistojen tuottamisesta ja jakelusta syntyy kustannuksia. Aineistokohtainen kustannusten ja tulojen erottelu edellyttäisi erittäin kehittyneitä ja läpinäkyvää kustannuslaskentaa. Keskustelun jatkamiseksi olisikin suositeltavaa, että tuotannon, omistamisen, jakelun ja tuottojen suhdetta selvitettäisiin käytännössä.

4.4 Data julkishyödykkeenä

Yksi lähestymistapa julkiseen dataan on ajatella sitä julkishyödykkeenä tai osana julkista kaikkien käytettävissä olevaa infrastruktuuria. Esimerkiksi tiet sekä taajamien katuvalot ovat kaikkien Suomessa vierailevien käytössä riippumatta siitä, maksaako veroja Suomeen vai ei. Julkishyödykkeellä tarkoitetaan sellaista hyödykettä, jonka käyttö ei ole poissulkevaa. Useat ihmiset voivat käyttää julkishyödykettä samanaikaisesti ilman että se kuluu tai sen käyttäjät estäisivät toisiaan. Kuronen (1998) käyttää esimerkkinä julkishyödykkeestä radio-ohjelmia, jotka tarjotaan kaikille kuunneltaviksi, eikä yhden kuuntelu haittaa muiden kuuntelunautintoa.

Julkisdatan taloudelliseen hallintaan liittyvä tarkastelu voidaan tämän oppaan kannalta jakaa kolmeen osaan: 1) tuottaminen, 2) omistaminen ja 3) käyttöön tarjoaminen ja rajakustannus. Markkina-ajattelussa nämä kolme tarkastelunäkökulmaa kietoutuvat tiiviisti toisiinsa. Aineiston tuottajasta tulee sen omistaja ja se voi tarjota aineistoa käytettäväksi maksua vastaan. Julkisen datan tapauksessa aineistojen tuottamisen rahoittaminen on syytä pitää erillään omistamiseen ja myymiseen liittyvistä kysymyksistä.

Tuottaminen

Julkishallinnon tietovarantojen kerääminen ei ole varsinaisesti markkinaehtoisia toimintaa, eikä datan ylläpidon jatkuminen yleensä ole riippuvainen siitä, pystytäänkö datasta luomaan myytäviä informaatiohyödykkeitä tai tehdäänkö sitä kannattavasti (Kuronen 1998). Osa aineistosta tuotetaan omaan käyttöön ja osaa tarjotaan muille. Toisaalta Suomessa on runsaasti aineistoja, esimerkiksi henkilödatan osalta, joissa samaa dataa kerätään turhaan päällekkäin moniin

rekistereihin. Joissakin tapauksissa taas esimerkiksi kunnat on lainsäädännöllä velvoitettuja keräämään raakadataa kansallisiin rekistereihin ilman, että itse hyötyvät kerätyistä datasta. Tällaisissa tapauksissa kunnat pyrkivät kompensoimaan lakivelvoitteesta syntyviä kustannuksia myymällä aineistoja.

Oletusarvoisesti julkishallinto siis kerää datavarantoja joka tapauksessa omaan käyttöön ja yleiseksi hyväksi, mutta maksuperustelain myötä syntynyt käytäntö on hämärtänyt tätä lähtökohtaa. Kyse on pitkälti näkökulmasta, kuten esimerkiksi tilanteessa, jossa kunta vuokraa itselleen omia kiinteistöjään. Budjetoinnin kannalta rahaa liikkuu eri yksiköiden välillä, mutta käytännössä rahan liikuttamisesta koituu vain kuluja.

Omistaminen

Hyvin usein informaation omistamiseen suhtaudutaan samalla tavoin kuin minikä tahansa muun omaisuuden omistamiseen. Edellisessä luvussa käsitelimme tekijänoikeuksia ja totesimme, että mikäli niitä luonnostaan syntyy, on niistä mahdollista luopua. Tieto itsessään ei ole muutoinkaan tekijänoikeuksien piirissä, vaan ainoastaan se muoto, missä se on esitetty. Jos julkishallinnon tuottamaa dataa ajatellaan julkishyödykkeenä, ei aineiston omistajuutta voida oikeastaan enää selkeästi määritellä. Aineiston tuottaminen ei siis välttämättä johda omistusoikeuden syntymiseen. Sen takia tässä oppaassa on puhuttu julkishallinnon hallussa olevista aineistoista eikä hallinnon omistamista aineistoista.

Käyttöön tarjoaminen ja rajakustannus

Kun data on koottu ja järjestetty tietokannoiksi, joista julkaistaan ladattavissa muodoissa tai rajapinnan kautta dataa käyttäjille, niin rajakustannukset yhdestä lisäkäyttäjistä ovat hyvin pienet tai olemattomat. Tämä ei poista sitä tosiseikkaa, että aineiston tuottaminen maksaa kuten myös alkuinvestointi toimivaan jakelujärjestelmään. Julkishallinnon data ei suinkaan ole ilmaista, mutta haluttaessa se voidaan luovuttaa maksutta käytettäväksi.

Kun datan keräämisestä ja julkaisuinfrastruktuurin ylläpidosta maksetaan korvaus sitä tekeville organisaatioille, niin syntyvä tulos eli raakadata voidaan jakaa maksutta julkishyödykkeenä. Kuten edellä todettiin, niin aineistojen ylläpito ja kerääminen tapahtuu yleensä muista syistä ja budjettivarojen. Jäljelle jää kysymys julkaisuinfrastruktuurin ylläpidosta eli siitä, tehdäänkö sekin budjettivarojen, vai peritäänkö aineistoista irtiottokustannuksiin perustuvia maksuja.

Ennen internetiä tiedon tuottaminen liittyi saumattomasti tiedon jakeluun, mutta ei enää nykyään. Internet mahdollistaa tehokkaan ja taloudellisen tavallaan levittää dataa julkishyödykkeenä. Julkishallinnon datan käyttöarvo ja hyöty saattaisi kasvaa merkittävästi suuremmaksi, jos se tarjottaisiin maksutta kaikkien käytettäväksi. Tasapuolisuuden ja yksinkertaisuuden nimissä dataa tulisi luovuttaa samanlaisin ehdoin julkishallinnon muille yksiköille, tavallisille kansalaisille ja yrityksille.

4.5 Aikakustannus

Mielenkiintoinen näkökulma avoimen datan taloudellisiin vaikutuksiin on ns. aikakustannus – toisin sanoen se, paljonko kansalaisilta ja virkamiehiltä kuluu aikaa informaation hankkimiseen ja kuinka moni potentiaalisesti hyödyllinen toiminto jää kokonaan tekemättä sen takia, että nyt informaation hankkiminen on liian työlästä. Aikakustannuksien avulla voidaan havainnollistaa tietovarantojen avaamisen suhdetta tuottavuuden lisääntymiseen. Innovatiivisuuden näkökulmasta rutiineihin ja mekaaniseen työskentelyyn käytetyn ajan vähentäminen vapauttaa resursseja luovaan ajatteluun ja toimintaan.

Demokratian ja kansalaisaktiivisuudenkin näkökulmasta ajan vapautumien on positiivista. Kansalaisilla on sitä paremmat mahdollisuudet osallistua yhteis-

kunnalliseen päätöksentekoon, mitä vähemmän aikaa sitoutuu pelkästään hallinnon prosessien seuraamiseen.

Aikakustannus on käyttökelpoinen näkökulma esimerkiksi vertailtaessa tilannetta, jossa jokin käyttäjälle tarpeellinen aineisto on ladattavissa suoraan internetistä siihen, että aineiston saamiseksi pitää ottaa yhteyttä viranomaiseen ja allekirjoittaa paperisopimus. Rahassa mitattuna sähköpostin lähettäminen ja lyhyt odottaminen saattaa vaikuttaa yhdentekevältä, mutta tieotyön maailmassa näin ei välttämättä ole. Asioiden hoitaminen yhdellä kertaa on edullista. Keskeytykset ja muiden toimijoiden odottelu aiheuttavat sen, että odottaja joutuu toistuvasti palauttamaan saman asian mieleensä ja tekemään ajatustyön uudelleen. Ihmiset toimivat useimmiten epätäydellisen tiedon varassa, sillä kaikkea relevanttia tietoa ei voida rajallisilla aikaresursseilla huomioida tai sen käyttöön-saaminen kestää liian kauan.

Myös julkishallinnon prosesseja voidaan mallintaa sen perusteella, kuinka paljon ne kuormittavat ajallisesti osallistujiaan. Hollannissa hallinto mittaa, kuinka paljon aikaa kansalaisilla vuositasona menee veroilmoituksen tekemiseen, hallinnon kirjeiden lukemiseen ja muuhun vuorovaikutukseen hallinnon kanssa (Den Hurk 2008). Tausta-ajatuksena on, että hallinnon kanssa asioimiseen käytetty aika on pois jostakin muusta. Näin saadaan mitattava suure, jonka avulla rinnakkaisia ratkaisuvaihtoehtoja voidaan arvioida ja pyrkiä minimoimaan hallinnon aiheuttamaa kuormitusta.

4.6 Iso-Britannian innovaatiovaluutta

Julkista dataa voidaan tarkastella myös innovointia sekä liiketoimintaa edistävänä katalyyttinä. Britannian Digital Britain -politiikkasuositukset (2009) puhuvatkin datasta *innovaatiovaluuttana*. Digital Britainin ajattelu jakaantuu kahteen suuntaan: Toisaalta on avointa ja julkista dataa, jota voidaan ajatella innovaatiovaluuttana. Se mahdollistaa ideoiden ja ajatusten vaihtamisen, joka jalostuu innovaatioiksi. Toisaalta ihmisten henkilökohtainen data on niinikään kasvavassa määrin käypää valuuttaa. Ihminen voi halutessaan esimerkiksi kertoa osan demografiatiedoistaan (kuten ikänsä tai sukupuolensa) vaihdossa eritasoisiin www-palveluihin, sosiaalisen pelaamisen pelipisteisiin, elokuvalippuihin, puheluminuutteihin jne. Yhtäältä Iso-Britannia haluaa siis vapauttaa julkisen datan vaihdantavälineeksi ja samalla suojella kansalaisten yksityistä dataa aiempaa paremmin.

Piilaakso on usein viitattu esimerkki ideoiden ja informaation avoimesta liikkuvuudesta ja sitä kautta verkottumisen merkityksestä. Toki Piilaakson menestyksen taustalla on muitakin tekijöitä, kuten se, että itse yhtiöt siirtyvät nopeammin ideasta tuotteeseen kuin mihin esimerkiksi Suomessa on vielä totuttu. Yksi merkittävimmistä tekijöistä on ideoiden nopea käyttöönotto. Ihmiset liikkuvat ongelmitta yhtiöstä toiseen ja kokonaisuus muodostaa yrittäjäekosysteemin sekä -verkoston. Tässä oppaassa esitetään synnyttäväksi avoimen datan ekosysteemiä hieman samalla mallilla.

Slush Helsinki on pari vuotta siten syntynyt ohjelmistoalan kasvuyrittäjien ja *born global* -yhtiöiden itse itselleen organisoima tapahtuma ja verkosto. Nämä yhtiöt jättävät kotimaiset markkinat väliin ja tähtäävät alusta alkaen globaaleille markkinoille, vaikka esimerkiksi TE-keskukset tarkastelevat kasvuyrityksiä paikallisen toimeliaisuuden näkökulmasta. Näillä yhtiöillä, kuten esimerkiksi peliteollisuudella, ei välttämättä edes ole riittäviä markkinoita Suomessa ja myös niiden rahoitus tulee ulkomailta. Slush on esimerkki Piilaakson kaltaisesta verkottuneesta keskittymästä.

Avoin data tarjoaisi muun muassa kasvuyrityksille raaka-aineistoa ja käytännön pilotointiympäristön globaaleja ohjelmistomarkkinoita varten. Lisäksi dataa voitaisiin hyödyntää esimerkiksi kaupallisten oppisisältöjen tuotannossa. Avoin data tukee myös Suomen uutta innovaatiostrategiaa (TEM 2008) yhtenä

katalysaattorina innovaatioverkostojen ja -ympäristöjen kehittymiseksi. Toisena esimerkkinä on pari vuotta sitten perustettu Sähköautot nyt! -yhteisö. Sen tavoitteena on luoda sekä ohjeet oman auton muuntamiseksi sähköautoksi ja edistää tarvittavaa infrastruktuuria Suomeen. Aluksi yhteisöä pidettiin erikoisena, mutta 2009 sen ehdotuksiin alettiin suhtautua vakavasti ja nyt yhteisön konseptista suunnitellaan jo vientituotetta Suomelle.

Toimittaja J. D. Lasica (2009) on koostanut raportin St. Aspenissa pidetystä kolmenkymmenen asiantuntijan aivomyrskystä. Työpajan aiheena oli tulevaisuuden hahmottaminen parinkymmenen vuoden aikajänteellä. Raportin yksi keskeisimpiä havaintoja on monen toiminnon siirtyminen pilveen eli toimivaksi kaikin puolin internetitse. Suomessa tätä ajattelua edustaa alueellisten teleyhtiöiden *Supermatrix*-hanke. Siinä käyttäjällä on päätelaite ja 100 Mbs laajakaista. Kaikki sovellukset, sisällöt, palvelut ja asetukset sijaitsevat pilvipalveluissa. Lasican raportti päättyi ajatukseen, että perinteiset valuutat alkavat ajan myötä korvautua internet-ympäristöissä paremmin toimiviksi vaihdantavälineiksi, joissa arvo määrittyy toiminnan eikä omistamisen kautta. Tällaisiksi nykyrahan korvaajiksi asiantuntijat nostivat muun muassa identiteetti- ja mainejärjestelmät, joita jo kokeillaan sosiaalisen median verkostopalveluissa, kuten Facebookin oheissovelluksena.

Avoim data osaltaan mahdollistaisi innovaatiostrategiamme mukaisesti Suomelle pilottiympäristön, jossa voitaisiin kokeilla myös uudenlaisia malleja datan tuotannon, omistamisen ja ylläpidon suhteen. Edelläkuvatut esimerkit osoittavat, että «ilmainen» eli maksuton voi olla kokonaisedullisesti paljon arvokkaampaa kuin nykyiset maksuperiaatteet.

4.7 Maksuttomuus kokonaisedullisuutena

Hallinnon tuottaman datan vapauttaminen on kansantaloudellisesti järkevää. Absoluuttisia tilastoja on vielä vähän. Eri alustavat laskelmat kuitenkin osoittavat, että kansalaisten, valtion ja liike-elämän kokonaishyöty on suurempi kuin nykyinen käytäntö ja maksuperustelaki. Hyötyjä tarkasteltaessa on hyvä huomata myös, kuinka paljon kuluja nykyinen malli aiheuttaa, kun esimerkiksi viranomaiset laskuttavat toisiaan, ja mitkä lopulta ovat tämän toiminnan tuotot? Kansantaloudellisten hyötyjen ohella julkisen datan avoimuus mahdollistaa myös muita toiminnallisia, yhteiskunnallisia ja sivistyksellisiä hyötyjä, joiden mittaaminen on ongelmallisempaa.

Esimerkiksi kattavassa MEPSIR (Measuring European Public Sector Information Resources) -tutkimuksessa Euroopan PSI-markkinoiden kooksi on arvioitu 27 mrd € (Dekkers & Al. 2006). Arvioit ovat vain suuntaa antavia ja etenkin arviot markkinoiden kasvupotentiaalista riippuvat muun muassa siitä, kuinka paljon oletetaan syntyvän kokonaan uutta liiketoimintaa ja lasketaanko mukaan vain suorat vaikutukset vai joitaikn epäsuoria taloudellisia vaikutuksia, kuten tuotavuuden lisääntyminen. Kaikesta huolimatta puhutaan suurista taloudellisista vaikutuksista.

Hallinnon avointa dataa pitäisi ajatella samanlaisena investointina kuin hallinnon muitakin työvälineitä laitteistoista ja ohjelmistoista sekä niiden ylläpidosta alkaen. Kun datan tuottamisen ja ylläpitämisen kustannuksista tehtäisiin arvio esiselvityksenä, summa voitaisiin sitoa valtion budjettiin esimerkiksi yleiskuluina. Tällöin datantuottajille maksettaisiin siitä, että he tekevät perustyötä datan keräämiseksi ja julkaisemiseksi. Analogia on yhteneväinen perustutkimukseen tai infrahankkeisiin ja niiden rahoittamiseen.

Verkossa toimii voimakas pienimmän yhteisen nimittäjän voima.
Monikäyttöiset, yksinkertaiset ja helposti omaksuttavat ratkaisut, joilla voidaan kattaa laajoja sovellusalueita, saattavat levitä yllättävänkin laajaan käyttöön.



5. Tekninen valmistautuminen

Tämä luku tarjoaa yleiskatsauksen avoimen datan tekniseen käsitteistöön. Näkemyksemme mukaan kaikkien datan avaamisen kanssa työskentelevien olisi hyvä ymmärtää avoimen datan teknologiset perusteet yleisellä tasolla. Toivomme teknologiaa selventävien määrittelyjen auttavan käytännön toiminnassa, kun sekä teknisen alan ammattilaiset että muiden alojen asiantuntijat ja päättäjät keskenään keskustelevat ja päättävät datan avaamisen politiikasta.

Aiemmissa luvuissa olemme opastaneet tekemään inventaarion organisaation tietovarannoista sekä tarkastelleet datan julkaisemista taloudellisesta ja juridisesta näkökulmasta. Nämä ovat toivon mukaan antaneet työkaluja päätöksentekoon. Seuraava vaihe on määrittellä datan julkaisemisen tekniset puitteet, joilla on suuri merkitys avoimen datan hyödyntämiselle ja jatkoprosessoinnille. Jos datan hyödyntäminen on liian vaikeaa, jäävät avoimen datan mahdollisuudet toteutumatta.

Kuten juridiikkaluvussa listasimme tärkeimmät avoimeen dataan liittyvät lait ja säännökset, niin myös tähän alkuun on koottu ja kuvailtu kaikki tässä luvussa esiintyvät teknologiat ja standardit (Taulukko 5.1). Standardien, yhteyskäytäntöjen eli protokollien ja formaattien (datan esitysmuotojen) määrää ei kannata säikähtää. Niiden muistaminen tai syvällinen sisäistäminen ei ole tarpeellista, sillä listoihin voi myöhemminkin palata. Tässä yhteydessä niitä käytetään pikemminkin apuneuvoina, joiden kautta voidaan viitata datan julkaisuun liittyviin teknisiin ratkaisuvaihtoehtoihin. Lista ei myöskään pyri olemaan kattava, mutta se sisältää näkemyksemme mukaan kuitenkin tärkeimmät teknologiat.

Taulukko 5.1 Avoimen datan julkaisuun liittyviä teknologioita

Standardeja datan esittämisen ohjelmallisesti käsiteltävissä muodossa	<p>XML: <i>Extensible Markup Language</i> (XML) on yleiskäyttöinen merkintäkieli, joka on laajennettavissa eri tarkoituksia varten uusilla merkintäelementeillä.</p> <p>CSV: <i>Comma Separated Value</i> (CSV) tiedostomuoto, jossa arvot erotetaan toisistaan pilkuilla. Tiedostot voidaan avata taulukkolaskentaohjelmilla.</p> <p>JSON: <i>JavaScript Object Notation</i> (JSON) on kevyt ohjelmointikielistä riippumaton tekstipohjainen datan siirtoformaatti.</p> <p>RDF: <i>Resource Description Framework</i> (RDF) on linked data-paradigman standardi, jossa yksittäisiä tietoresursseja kuvaillaan niihin linkitettävien sanastojen avulla. (Tässä luvussa RDF-lyhenteellä viitataan XML-muotoisiin RDF-tiedostoihin, joiden nimeämisessä monasti käytetään .rdf-päätettä.)</p>
XML-pohjaisia eri tarkoituksiin kehitettyjä erikoistuneita merkintäkieliä	<p>RSS ja GeoRSS: <i>Really Simple Syndication</i> (RSS) on XML-pohjainen tietoformaatti päivittyvien verkkosisältöjen eli syötteiden välitykseen. Mikäli syötteeseen liitetään paikkakoordinaatit, on kysessä GeoRSS-syöte.</p> <p>ATOM: Atom nimi viittaa kahteen toisiaan lähellä olevaan standardiin. <i>Atom Syndication Format</i> on XML-pohjainen merkintäkieli verkkosyötteiden (vrt. RSS) esittämiseen ja <i>Atom Publishing Protocol</i> (AtomPub) on yksinkertainen HTTP-protokollaan pohjautuva ohjelmointirajapinnan kuvaava standardi blogien päivittämiseksi.</p> <p>KML: <i>Keyhole Markup Language</i> (KML) on XML-pohjainen merkintäkieli paikkatietojen merkitsemiseen ja niiden näyttämiseen karttapalveluissa.</p>
Asiakirjojen esittämiseen tarkoitettuja tiedostomuotoja	<p>HTML: <i>Hypertext Markup Language</i> (HTML) on www:n keskeisin tiedostomuoto, joka mahdollistaa verkkosivujen rakenteen (mutta ei sisällön rakenteen) esittämisen ja sivujen linkittämisen toisiinsa yhdeksi hypertekstiverkoksi.</p> <p>PDF: <i>Portable Document Format</i> (PDF) internetissä yleinen dokumentin hyvää tulostettavuutta painottava tiedostomuoto.</p>
Siirtoprotokollia	<p>HTTP: <i>Hypertext Transfer Protocol</i> (HTTP), eli hypertekstin siirtoprotokolla on yksi internetin keskeisimmistä standardeista.</p> <p>PuSH: <i>PubSubHubbub</i> (PuSH) siirtoprotokolla nopeaan tietopäivityksistä ilmoittamiseen.</p>
Tekniikoita metatietojen sisällyttämiseen verkkosivuille	<p>Mikroformaattit ovat joukko pieniä HTML:n elementeistä muodostettuja formaatteja koneluettavan tiedon upottamiseksi www-sivuihin.</p> <p>RDFa: formaatti koneluettavien merkityksien upottamiseksi www-sivuihin.</p>
Resurssien nimeäminen	<p>URI: <i>Universal Resource Identifier</i> (URI) on internet-resurssin tunniste</p>
Arkkitehtuurityyli-rajapintojen tekemiseen	<p>REST: <i>Representational State Transfer</i> (REST) on HTTP-protokollaan perustuva arkkitehtuurimalli ohjelmointirajapintojen toteuttamiseen.</p>

5.1 Suunnittelukysymyksiä

Datan koneluettavuus ei ole suoraviivainen mekaaninen toimenpide. Seuraavassa esitämme muutamia tyypillisiä eri toteutusvaihtoehtoihin liittyviä kysymyksiä, jotka nousevat esille siinä vaiheessa, kun jokin aineisto päätetään julkaista verkkoon:

- **Missä formaatissa julkaistaan?**
- **Mitä standardeja käytetään?**
- **Miten julkaistava tieto pitäisi kuvailla ja mitä metadataa tarjota?**
- **Pitäisikö tehdä rajapinta vai tarjota aineisto ladattavaksi tiedostomuodossa?**

Avoin lähdekoodi, avoin data ja avoimet formaatit ovat tietohallinnon strategisia työkaluja, joiden avulla voidaan välttää riippuvuus yksittäisestä toimijasta. Julkisen hallinnon tietohallinnon neuvottelukunta (JUHTA) julkaisi vuonna 2009 suosituksen avoimen lähdekoodin ohjelmien käytöstä julkisessa hallinnossa (JHS 169). Siinä missä avoin lähdekoodi mahdollistaa sen, että hallinnon ohjelmistokehitys ei ole yhdestä toimittajasta riippuvaista, avoin data mahdollistaa sen, ettei hallinnolla ole tahatonta monopolia uusien palveluiden ideointiin ja kehittämiseen. Luonnollisesti avoimen ja myös suljetun datan siirrossa tarvitaan alusta- ja ohjelmistoriippumattomia siirtoformaatteja.

Standardeja tarvitaan. Avoimet standardit ovat edellytyksiä avoimille markkinoille. Vaikkei standardeja arkisesti tule ajatelleeksi, niin ilman niitä tuotteen ostajat olisivat aina riippuvaisia valmistajasta eikä jo hankittuihin tuotteisiin voitaisi hankkia lisäosia tai muuta yhteentoimivaa tuotetta. Jokainen meistä pitää itsestäänselvänä, että ostamamme energiansäästölamppu sopii aikaisemmin omistamaamme valaisimeen. Tämä ajattelumalli ei ole vielä kovin vanhaa tietotekniikka-alalla. Esimerkiksi vasta hiljattain on saatu EU:n avustamana matkapuhelinvalmistajat sopimaan siitä, miten matkapuhelimen laturi liitetään puhelimeen. Tulevaisuudessa matkapuhelimen ostaja ei enää ole riippuvainen puhelimen valmistajasta, mikäli laturi häviää ja hänen pitää ostaa tai lainata sitä (EU 2009).

Avoin standardointi ei järjestelmätoimittajan tärkeysjärjestyksessä aina ole kovinkaan korkealla. Entuudestaan vahvoilla oleva valmistaja hyötyy siitä, että asiakkaat ovat riippuvaisia sen tuotteista esimerkiksi valmistajan standardoimattomien tiedostomuotojen käytön yleisyyden vuoksi. Kilpailutilanteen muuttuessa vahvoilla ollut toimittaja saattaa myöhemmin tehdä myönnytyksiä ja julkistaa standardin muiden käytettäväksi, mutta pitää edelleen standardin kehitystyön itsellään tai perii maksua standardin käytöstä.

5.2 Mitä on koneluettavuus

Merkintäkielenä HTML (taulukko 5.1) soveltuu hyvin asiakirjojen rakenteen kuvailuun (mikä on otsikkoa ja mikä leipätekstiä), mutta itse tietosisältöä se ei tarjoa koneluettavassa muodossa. Esimerkiksi jos kunnan verkkosivuilla luetellaan kaikkien kirjastojen osoitteet ja aukioloajat on ihmisen helppo löytää haluamansa tieto sivuilta, mutta koneellisesti on vaikea erottaa osoitteita, aukioloaikoja ja muuta sivuilta löytyvää informaatiota toisistaan.

Jos asiaa katsotaan hetken kuvitteellisen «Kunnallispalvelut kännykässä» mobiilisovellusta kehittävän ohjelmoijan näkökulmasta, joutuisi hän pahimmillaan kopioimaan aukiolotiedot käsin ohjelmaansa internet-sivulta - tiedonmuruksia toiseen. Ensimmäinen askel ohjelmoijan avustamiseksi olisi tarjota osoite- ja aukiolotiedot pilkku- tai sarkainerotellussa tekstitiedostossa (esim. CSV) ladattaviksi. Tällöin ohjelmoija voisi kerralla lukea kaikki kirjastojen osoitteet ja aukioloajat ohjelmaansa koneellisesti.

Yksinkertaistetussa esimerkissämme ohjelmoijan seuraava haaste on eri lähteistä kerättyjen kirjastojen, päiväkotien ja uimahallien osoite- ja aukioloaikatietoja kuvaavien aineistojen yhdistäminen. Koneluettavuuden näkökulmasta

on pystyttävä yhdistämään eri aineistojen metatiedot, jotta koneellisesti voidaan todeta, kirjastojen «osoitteiden» olevan verrannollisia «katuosoitteeseen» uimahalleista puhuttaessa. Aina tietueiden yhdistäminen (harmonisoiminen) ei kuitenkaan ole näin suoraviivaista. Hyvän esimerkin tarjoavat työttömyysluvut, jotka Tilastokeskus ja työ- ja elinkeinoministeriö laskevat eri tavalla. Näitä lukuja ei voida helposti yhdistää.

Pilkku- tai sarkaineroteltu data sopii parhaiten taulukkomuotoisen tiedon tai nimi-arvoparien julkaisemiseen. Etuna on, että lähes jokaiselta tietokoneelta löytyy taulukkolaskentaohjelma, jolla tällaiset tiedostot voidaan avata tutustuttaviksi ja muokattaviksi. Monimutkaisempien tietorakenteiden esittämiseen ainakin XML, JSON ja RDF ovat mainitsemisen arvoisia yleisen tason standardeja.

Koneiden välisessä tiedonsiirrossa ja automaattisessa prosessoinnissa verkon «lingua franca» on XML. Se ei ole valmis formaatti tiedon julkaisemiseen, vaan standardi, jolla voidaan määritellä sovelluskohtaisia merkkaukieliä tietojen esittämiseksi. Monimutkaisten tietorakenteiden julkaisemiseen voidaan käyttää jotakin olemassa olevaa XML:ään perustuvaa kieltä tai määritellä oma XML-rakenne. Eri merkkaukielien harmonisoimiseksi linked data -paradigmassa tietueet esitetään RDF-muodossa (kohta 5.6).

JSON on puolestaan XML:ää kevyempi tapa esittää ohjelmallisesti käsiteltävää dataa mahdollisimman helposti siirrettävässä muodossa. JSON-syntaksin avulla voidaan esittää sekä yksinkertaisia nimi-arvopareja että monimutkaisia tietorakenteita siten, että niitä on helppo käsitellä paljon käytettyjen internet-ohjelmointikielten avulla.

5.2.1 Rajapinnat ja esitysmuodot datan mukaisesti

Kun aineistojen avaamisesta on päätetty, on usein ensimmäisiä kysymyksiä, missä formaatissa dataa pitäisi jakaa. Tiedon käyttäjille tärkeintä on, että data on saatavilla ohjelmallisesti käsiteltävässä muodossa ja esitysmuoto on avoin, eli että sen käsittely ei edellytä minkään tietyn toimittajan ohjelmistoja. Esitysmuodon avoimuus ja koneluettavuus ovat minimivaatimukset, jotta dataa päästään koneellisesti käsittelemään. Esitysmuodot liittyvät datan käytön helppouteen (tai vaivalloisuuteen) myös siten, että jokainen muunnos muodosta toiseen vaatii työtä; jotkut muunnokset ovat helpompia kuin toiset.

Pilkku- tai sarkaineroteltu data sopii parhaiten taulukkomuotoisen tiedon tai nimi-arvoparien julkaisemiseen. Etuna on, että lähes jokaiselta tietokoneelta löytyy taulukkolaskentaohjelma, jolla tällaiset tiedostot voidaan avata tutustuttaviksi ja muokattaviksi. Monimutkaisempien tietorakenteiden esittämiseen ainakin XML, JSON ja RDF ovat mainitsemisen arvoisia yleisen tason standardeja.

Koneiden välisessä tiedonsiirrossa ja automaattisessa prosessoinnissa verkon «lingua franca» on XML. Se ei ole valmis formaatti tiedon julkaisemiseen, vaan standardi, jolla voidaan määritellä sovelluskohtaisia merkkaukieliä tietojen esittämiseksi. Monimutkaisten tietorakenteiden julkaisemiseen voidaan käyttää jotakin olemassa olevaa XML:ään perustuvaa kieltä (esim. paikkatietojen esittämiseen käytetty KML) tai määritellä oma XML-rakenne. Eri merkkaukielien harmonisoimiseksi linked data -paradigmassa tietueet esitetään RDF-muodossa (kohta 5.6).

JSON on puolestaan XML:ää kevyempi tapa esittää ohjelmallisesti käsiteltävää dataa mahdollisimman helposti siirrettävässä muodossa. JSON-syntaksin avulla voidaan esittää sekä yksinkertaisia nimi-arvopareja että monimutkaisia tietorakenteita siten, että niitä on helppo käsitellä paljon käytettyjen internet-ohjelmointikielten avulla.

Ensivaiheessa datan tarjoajan kannattaa saattaa data päivänvaloon itselleen

helpoimmalla tavalla. Myöhemmin datan käytettävyyttä voidaan lisätä tekeillä muunnoksia muihin esitysmuotoihin. Nyrkkisääntönä voidaan pitää, että muunnokset yleisesti käytettyihin esitysmuotoihin kannattaa tehdä datan tarjoajan päässä sen sijaan, että jokaisen datan hyödyntäjän pitäisi tehdä sama muunnos erikseen. Sama aineisto voidaan tarjota esimerkiksi XML-, JSON- ja RDF-muodossa (kohta 5.5).

Kun data on koneluettavasti saatavilla, se on dokumentoitava, jolloin voidaan kertoa mitä mikäkin tieto tarkoittaa. Esimerkiksi pilkku- tai sarkaineroteltuna voidaan koneluettavasti esittää taulukkomuotoista tietoa, mutta se ei ota kantaa siihen, mitä sarakkeita taulukossa on. Ilman sarakeotsikoita moni taulukko olisi melko hankalasti käsiteltävissä. Siksi koneellisesti käsiteltävän muodon ohella on tärkeää tarjota dokumentaatio siitä, mitä rakenne pitää sisällään. Esimerkitapauksessa osoitteista ja aukioloajoista voi sisällöstä päätellä merkityksen, mutta yleensä tapaukset ovat monimutkaisempia ja selkeä tiedon käyttööpas on tarpeen.

5.2.2 Koneluettavat lisenssit

Samalla kun huolehditaan datan tarjoamisesta koneluettavassa muodossa, on hyvä tarjota myös aineiston käyttölisenssi samalla tavalla. Lisenssien koneluettavuus helpottaa yhdistelmäpalveluiden kehittäjien mahdollisuuksia seurata ehtojen noudattamista, sillä käyttöehdot voidaan huomioida palvelun toiminoissa.

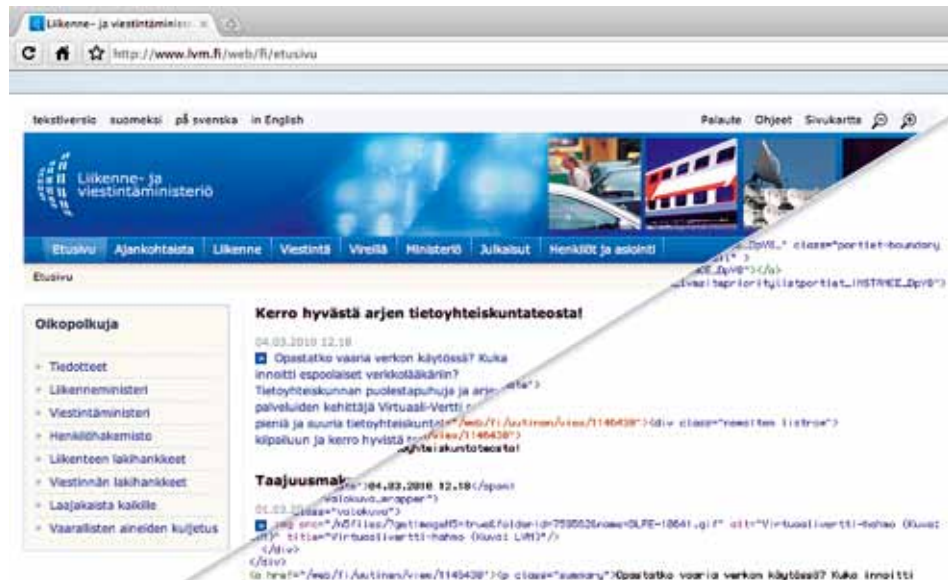
Koneluettavuus parantaa myös aineistojen löydettävyyttä. Esimerkiksi Flickr-kuvapalvelusta on mahdollista hakea hakusanaa vastaavia kuvia, joiden lisenssiehdot sallivat niiden uudelleenkäytön vaikkapa omassa powerpoint-esityksessä. *Creative Commons* -lisensseillä (Liite 4) on ollut mahdollista ilmaista käyttöoikeudet koneluettavasti jo vuosikausia. Vaikka aineisto olisi julkishyödyke, on sekin syytä ilmaista eksplisiittisesti. Tämä tarve vain korostuu koneluettavuuden myötä. *Creative Commons Waiver* -lisenssi (CC0) sopii tähän tarkoitukseen. Myös PSI-direktiivissä suositellaan lisenssien tarjoamista koneluettavassa muodossa.

5.3 Web-arkkitehtuuri

Termejä *www* ja *internet* käytetään usein erheellisesti toistensa synonyymeinä. Wwv on kuitenkin vain yksi monista internetin palvelumuodoista. Muita palvelumuotoja ovat esimerkiksi sähköposti, keskusteluryhmät tai tiedoston siirto. Internetille luotiin pohjaa jo 1960-luvulla Yhdysvalloissa, ja nykyisin se on maailmanlaajuinen eri tietoverkkoja yhdistävä laitteisto- ja ohjelmistoinfrastruktuuri, jonka avulla tietokoneet voivat olla yhteydessä toisiinsa. Tim Berners-Leen 1990-luvulla kehittämä *www* on puolestaan internet-verkossa toimiva hajautettu linkkeihin perustuva hypertekstijärjestelmä. Kun tässä oppaassa puhutaan tietovarantojen julkaisemisesta avoimesti verkkoon, tarkoitetaan sillä sitä, että ne tuodaan osaksi *www*:tä ja web-arkkitehtuuria.

W3C-standardointiorganisaatio edistää *www*:n kehittymistä rakentaen yhteisiä sopimuksia edistämään webin eri osien yhteensopivuutta. W3C sai vuonna 2004 valmiiksi suosituksen, joka kirjaa kokonaisuutena toimivan web-arkkitehtuurin keskeiset osat ja suunnitteluperiaatteet (W3C 2004). Valitettavasti sen suosituksia ei aina tunneta tai noudateta esimerkiksi ohjelmallisesti luotujen URI-tunnuksien muodostamisen osalta. Suositus tarjoaa hyvän kehikon datan tuomiselle verkkoon.

Web-arkkitehtuuri mahdollistaa asteittaisen kehityksen hajautetussa verkko-ympäristössä. Se on osoittanut muuntautumiskykyä ja yhden organisaation kas-



Kuva 5.1: Hypertekstiä luetaan selaimella, joka hakee verkkosivuiksi kutsuttuja html-dokumentteja verkkopalvelimilta ja esittää ne ihmisille luettavassa muodossa.

vavan linkittyneen informaatioavaruuden voiman. Myöhemmin tässä luvussa käsiteltävä linkitetty data esittelee, miten dataa voidaan julkaista internetissä kytkeytyen osaksi maailmanlaajuista www-informaatioavaruutta. Samoin rajapintojen yhteydessä esiteltävät REST-tyylin palvelurajapinnat perustuvat web-arkkitehtuuriin ja soveltuvat siksi hyvin datan julkaisemiseen avoimessa verkossa.

Verkossa toimii voimakas pienimmän yhteisen nimittäjän voima. Monikäyttöiset, yksinkertaiset ja helposti omaksuttavat ratkaisut, joilla voidaan kattaa laajoja sovellusalueita, saattavat levitä yllättävänkin laajaan käyttöön. Aikaa ja vaivaa säästyy, kun teknologia-komponentteja, osaamista ja käytäntöjä voidaan soveltaa ongelmien ratkaisemiseen. Näin on käynyt www:lle, kun siitä on tullut yleinen väline erilaisten palveluiden toteuttamiseen. Käytännössä melkein koko päivittäin käyttämämme internet perustuu muutamaankin erittäin laajasti omaksuttuun standardiin (TCP/IP, HTTP, HTML, CSS, JavaScript jne.). Nämä standardit ovat tuttuja kehittäjille, ylläpitäjille, ohjelmistoarkkitehdeille ja web-arkkitehtuuriin perustuvien järjestelmien toteuttamiseen löytyy paljon valmiita avoimen lähdekoodin ohjelmistoja ja edistyskäsittelyä työkaluja.

Yksi tärkeimmistä työkaluista on päivittäin käyttämämme selain (kuva 5.1). Alunperin hypertekstin selaamiseen tarkoitettujen selainohjelmien ominaisuudet ovat laajentuneet ja niitä käytetään www:n ohella käyttöliittymänä moniin muihinkin internetin palvelumuotoihin, kuten edellä mainittuihin sähköpostiin ja tiedonsiirtoon, sekä pikaviestintään sosiaalisen median palveluissa. Www-terminikin alkaa hävitä puhekielestä.

Web-arkkitehtuurin keskeisin käsite on globaali informaatioavaruus, joka koostuu yksiselitteisesti tunnistettavista toisiinsa linkitetyistä resursseista. Resurssi voi esimerkiksi olla dokumentti tai verkkoon kytketty tietokoneohjelma, jolla on yksikäsitteinen tunnistus. Kaikille verkon käyttäjille tuttuja URI-tunnuksia ovat internetosoitteet, kuten <http://www.suomi.fi>. Web-arkkitehtuurin piiriin tuotavien asioiden nimeämiseen suositellaan käytettäväksi HTTP-URI:ia muiden merkintätapojen sijaan.

Sama resurssi - vaikkapa tieto eduskunnan täysistunnon äänestystuloksista - voidaan esittää useissa esitysmuodoissa, kuten HTML, CSV, XML tai RDF (ks 5.4). Web-arkkitehtuurin ytimessä on hypertekstin siirtoprotokolla HTTP (Hypertext Transfer Protocol), joka määrittelee mahdolliset toiminnot (GET, PUT, POST ja

DELETE) selaimien ja www-palvelimien väliseen vuorovaikutukseen. Tämä perusrakenne on osoittautunut toimivaksi sekä verkossa surffaavien ihmisten että tietokoneohjelmien ohjelmoimisen näkökulmasta.

5.4 Sisällön esitysmuotoja

Internetsovelluksissa käytetään IANA:n (Internet Assigned Numbers Authority) standardoimia sisällön esitysmuotoja eli niin sanottuja MIME-mediatyyppejä. Esitysmuotoihin on listattu yleisimpien tietokoneohjelmien tiedostomuodot sekä vain internetissä esiintyviä esitysmuotoja. Avoimien esitysmuotojen käsitteilyyn on olemassa paljon ohjelmistokirjastoja ja työkaluja eri ohjelmointikielille. Jotkin valmistajakohtaiset esitysmuodot ovat myös yleisiä internetissä, kuten Microsoft Excel-tiedostomuodot. Näiden käyttö on sidotumpaa valmistajan omiin työkaluihin.

5.4.1 Ilmoitukset muutoksista syötteiden avulla

Usein päivittyvät tiedot voidaan esittää käyttäjälle tietovirtana, josta näkee ajantasaisesti tuoreimmat muutokset. Paras tapa ilmoittaa muutoksista on julkaista syöte. Syötteen tilaajat saavat automaattisesti tiedon muutoksesta omaan järjestelmäänsä (esim. syötteiden lukijaan, selaimeen tai vaikka sähköpostiohjelmaan). Syötteissä julkaistavia tietoja on myös melko helppo hyödyntää ohjelmallisesti, joten joissakin tapauksissa ne kelpaavat sellaisenaan datan jakamiseen. Syötteiden ilmaisuvoima koneluettavien merkityksien ilmaisemiseen on kuitenkin heikko. Muutoksista ilmoittamisen ohella syötteet ovat hyvä tapa välittää otsikoita, lyhyehköjä tekstejä ja asiaan liittyviä linkkejä, kuten Hovioikeuksien ratkaisujen tapauksessa on tehty (<http://www.oikeus.fi/rss/ho/hovioikeuksienratkaisut.rss> - löytyy suomi.fi/dataopas).

Syötteiden kuvaamiseen on olemassa kaksi yleisesti käytettyä esitysmuotoa: RSS ja Atom. Teknisesti RSS ja ATOM ovat XML-merkintäkieliä. Jakamalla tuottamansa tiedon syötetiedostomuodossa sisällöntuottaja mahdollistaa tiedon uudelleenjulkaisemisen muissa verkkopalveluissa ja tiedon säännöllisen seuraamisen syötteenlukuohjelmilla. RSS-muotoa käytetään nykyisin erittäin laajasti mm. uutisten, blogien ja muun ajankohtaisen sisällön, kuten tiedotteiden, välittämiseen.

5.4.2 Ajantasainen verkko

Uusien teknologioiden ansiosta dataa voidaan tuottaa ja seurata maailmanlaajuisesti ja ajantasaisesti. Ajantasaisen datan jakelu ja seuraaminen (real-time web) on nopeasti lisännyt merkitystään internetin käytössä. Miten sitten julkihallinnon data liittyy reaaliaikaiseen internetiin? Terveyskeskusten jonotustilanne, junien pysyminen aikataulussa ja säätiedot ovat esimerkkejä sovelluskohteista, joissa ajantasainen data helpottaisi ihmisten elämää. Viivästyneenä saman datan käyttömahdollisuudet ovat rajallisempia. Suurin hyöty datasta saadaan julkaisemalla se sekä ajantasaisesti että arkistoimalla se myöhempää käyttöä varten. Tällöin dataa voidaan jälkikäteen käyttää esimerkiksi toiminnan kehittämiseen.

Ongelmana ajantasaisuuteen pyrkimisessä on se, että internetiä ei ole alunperin suunniteltu ajantasaisuutta varten. Tähän asti verkkosovellukset toimivat uuden informaation saamiseksi kyselyperiaatteella lähettäen julkaisijalle jatku-

vasti «onko mitään uutta» -kyselyitä, vaikka 99 % tapauksista vastaus on «ei». Ajantasaisten palveluiden tekemiseen on vasta hiljattain yleistynyt protokollia, joilla tämä toimintalogiikka on käännetty toisinpäin eli tiedon julkaisijataho lähettää halukkaille tilaajille uuden informaation silloin, kun sitä on. Tällä hetkellä voimakkaimmin suositetaan kasvattava protokolla ajantasaisessa internetissä on PubSubHubbub (PuSH), jonka avulla uusi sisältö lähetetään suoraan tilaajalle.

5.4.3 Paikkatieto uudelleenkäytettävässä muodossa

Moni tieto liittyy tiettyyn fyysiseen sijaintiin, kuten paikkaan, alueeseen tai osoitteeseen. Paikkatieto laajasti käsitettynä kattaakin merkittävän osan julkishallinnon tuottamasta informaatiosta, sillä sijaintitieto yleensä koordinaattien, aluetunnusten tai osoitteiden muodossa liittyy moniin hallinnon ylläpitämiin rekistereihin. Paikkatiedon erityispiirteenä on, että sen hallinnointi on ollut perinteisesti varsin ammattimaista ja verrattain teknisorientoituneen henkilökunnan käsissä, minkä ansiosta sen uudelleenkäytettävyyden ja jakamisen eteen on tehty paljon enemmän töitä kuin esimerkiksi tekstimuotoisen asiakirjatiedon.

Jos tieto sijainnista on koneellisesti käsiteltävässä muodossa, niin tietoja voidaan esittää havainnollisesti kartalla ja eri lähteiden tietoja voidaan yhdistellä toisiinsa niiden sijainnin perusteella. Nykyisin karttapalveluiden teko on ohjelmoijille melko yksinkertaista olemassa olevien rakenteiden, kuten Open Streetmap tai Google Maps, avulla. Paikkatiedon uudelleenkäyttö on yleistyneiden karttasovellusten muodossa helposti ymmärrettävää, mikä on vilkastuttanut sen ympärillä käytyä keskustelua.

Taulukko 5.2. Esimerkkejä tiedoista, joihin tyypillisesti liittyy sijainti

Päätöksenteko	kaavamuutosehdotukset, tiettyä aluetta tai osoitetta koskevat esitykset ja päätökset,...
Palvelut	kaupalliset ja julkiset palvelut; kirjastot, sairaalat ja liikuntapaikat aukioloaikoineen; kaupat, ravintolat, elämyspalvelut,...
Liikenne	ruuhkatiedot, tietyt, julkisen liikenteen pysäkit aikatauluineen sekä poikkeusliikennetiedotukset vaikutusalueineen, ajoneuvojen ajantasaiset sijainnit,...
Säätiedot	tiesää, sääennusteet, ...
Mediasisältö	valokuvat, videot, johonkin paikkaan liittyvät tarinat,...

Teknisesti sijaintitiedon lisääminen esimerkiksi RSS-syötteenä julkaistaviin tietoihin on melko yksinkertaista. Sijaintimerkinnät sisältävää syötettä kutsutaan GeoRSS-syötteenä. Google Maps -ohjelmointirajapinnassa on valmiit toiminnot GeoRSS-syötteen tietojen esittämiseen kartalla. Vielä monipuolisemmin sijaintitietoa voidaan julkaista KML Network Link -toiminnallisuuden avulla.

Google Maps ja Google Earth osaavat hyödyntää KML-syötteitä niin, että käyttäjälle haetaan tietoja aina näkyvissä olevalta alueelta ja tietoja päivitetään näkymän muuttuessa. Tästä syystä syötteenä kannattaa julkaista myös sellaisia tietoja, jotka eivät itsestään päivyty, mutta joista tulee käyttäjälle merkityksellisiä, kun käyttäjä tarkastelee tiettyä aluetta. Samaa GeoRSS- tai KML-syötettä voidaan hyödyntää useassa karttapalvelussa (tai muussa sijaintitietoa hyödyn-

tävässä palvelussa). Toisaalta yhteen palveluun voidaan liittää GeoRSS- ja KML-syötteitä useista lähteistä.

5.4.4 Asiakirjojen julkaiseminen

Suomen kielen perussanakirjan mukaan asiakirja on «määrätarkoitukseen käytävä kirjallinen esitys». Asiakirjoja voidaan toki säilyttää myös digitaalisessa muodossa, mutta ajatuksellisesti sana viittaa usein kiinteisiin artefakteihin, kuten sopimuspaperit tai passi. Vähintäänkin vaaditaan, että digitaalinen asiakirja pitää voida tulostaa. Tyypiesimerkki nykyään täysin tarpeettomasta asiakirjasta on lentolippu. Matkatoimisto lähettää sen sähköpostitse ja matkustaja tulostaa varmuuden vuoksi. Monissa tapauksissa olisikin perustellumpaa julkaista ja välittää asiakirjojen sisältämä informaatio sellaisenaan.

Kaupunginvaltuuston kokousten asialista on tyypiesimerkki asiakirjoista. Nykyisin kaupungin hallintokoneistoissa erilaiset päätösasiat etenevät aloitteista kokouksien kautta lausunnoille, lautakuntiin ja taas uusiin kokouksiin. Päätösasioita hallinnoidaan usein erillisillä asianhallintajärjestelmillä, joista sitten yksittäisen kokouksen asialistat julkaistaan ennen kokousta internetiin. Yksittäisen asian etenemisestä kiinnostunut joutuu helposti käymään läpi ison pinon pitkiä asialista- ja päätöspöytäkirjadokumentteja löytääkseen etsimänsä. Mikäli hänellä olisi sen sijaan mahdollisuus tehdä hakuja asianhallintajärjestelmän sisäiseen tietokantaan, saisi hän yhdellä tai muutamalla haullla koottua itselleen merkityksellisen informaation.

5.5 Rajapinnat, sovellukset ja palvelut

Mitä tarkoitetaan, kun puhutaan avoimista rajapinnoista? Tietokoneohjelmien käyttöliittymä on ihmiskäyttäjän ja ohjelman välinen rajapinta, jonka kautta kommunikaatio tapahtuu ja informaatio liikkuu. Vastaavasti ohjelmien välillä on rajapintoja, joiden kautta ohjelmat tai ohjelmiston osat kommunikoiivat keskenään. Usein rajapinnat on tarkoitettu vain järjestelmän sisäiseen käyttöön ja mahdollisiin ennalta sovittuihin tietotekniikkajärjestelmien integrointeihin. Nykyisin on kuitenkin erittäin yleistä toteuttaa esimerkiksi verkkopalveluun avoin ns. web API (application program interface) eli verkko-ohjelmointirajapinta, jota on mahdollista käyttää internetin välityksellä.

Tyypillisesti web API on toteutettu järjestelmään, joka voi tarjota erilaisia valmiita palveluita muille ns. sovellusohjelmille. Tämä helpottaa ohjelmointityötä, kun kaikkea ei tarvitse tehdä moneen kertaan. Ohjelmoiden suosima Googlen karttarajapinta tarjoaa esimerkiksi kätevän geokoodauspalvelun, joka muuntaa tekstimuotoiseen osoitteen (ja paljon muutakin) vastaavaksi maantieteelliseksi koordinaatiksi. Tällaisen ohjelmoiminen itse olisi työlästä, vaikka kaikki tarvittava data olisikin käytössä.

Rajapinnan kautta tarjottava palvelu voi olla yksinkertaisimmillaan sellainen, että lähettämällä sopivan pyynnön palvelu vastaa palauttamalla tiedoston. Näin yksinkertaista palvelua ei mielletä välttämättä palveluksi, joten puhutaan pelkistä rajapinnoista, joiden kautta saa dataa. (Tarkoin termein tällaisia kutsutaan tietopalvelurajapinnoiksi.)

Palvelurajapinnat on yksi tapa tarjota dataa koneluettavassa muodossa tietoteknisten järjestelmien käytettäväksi. Esimerkkejä palveluista, jotka tarjoavat alunperin julkishallinnon tuottamaa dataa rajapintojen kautta, ovat Google Transit API sekä brittiläisen Guardian-lehden World Government Data API. Google Transit API tarjoaa julkisen liikenteen reittitietoja maailmanlaajuisesti ja Guardianin rajapinnan kautta käsiksi tällä hetkellä Iso-Britannian, USA:n, Aust-

ralian, Uuden Seelannin ja Lontoon julkisen datan katalogien sisältöön.

Ohjelmille rajapintojen kautta tarjottavia palveluita ei pidä sekoittaa HTML:ään perustuviin ihmisten luettavaksi ja käytettäväksi tarkoitettuihin verkkopalveluihin. Esimerkiksi Helsingin seudun liikennelaitoksen reittiopas (ennen YTV:n reittiopas) on yksi Suomen suosituimmista verkkopalveluista, mutta reittiopasjärjestelmään on olemassa myös ohjelmointirajapinta, jonka kautta voidaan ohjelmallisesti hakea muun muassa aikataulu- ja pysäkkietietoja. Tällöin reittiopasjärjestelmä tarjoaa datan rajapinnan kautta palveluna muille sovellusohjelmille, kuten vaikkapa iPhone-kännykkäsovellukseen, jonka kautta reittejä voi suunnitella. Liitteessä 5 on käsitelty tarkemmin yleinen palvelukeskeinen arkkitehtuuri (SOA Service Oriented Architecture) ja ns. REST -tyylin rajapinnat, jotka ovat suosittuja avoimen datan jakamiseen.

5.6 Linkitetty data

Toisiinsa linkitettyihin resursseihin perustuva web-arkkitehtuuri on maailmanlaajuinen tietoarkkitehtuuri, johon avoimesti dataa julkaisevien tahojen tulisi datavarantonsa liittää, jotta ne olisivat tehokkaimmin kaikkien käytettävissä. Kun aikanaan ihmiset alkoivat spontaanisti solmia yhteen HTML-sivuja linkkien ja URL-osoitteiden avulla, muodostui internet sellaiseksi kuin me sen nykyään arkisesti tunnemme, ts. verkoksi toisiinsa linkitettyjä dokumentteja. Vastaavalla tavalla on mahdollista ja nykyisin jo yleistä linkittää yhteen HTML-sivujen ohella myös dataa. Kehitys on synnyttämässä uudenlaista internetiä, verkkoa toisiinsa linkitettyä dataa. Termillä linkitetty data (Linked Data) viitataan yleensä [www:n](http://www.w3.org/) kehittäjän Tim Berners-Leen vuonna 2006 julkaisemaan neljään periaatteeseen (Taulukko 5.3).

Taulukko 5.3: Linkitetyn datan neljä periaatetta (Berners-Lee 2006)

Teknisemmin:	Tarkoituksena:
1. Käytä URI:ja (Universal Resource Identifier) niminä asioille	1. muodostaa asiasta käsite, josta voidaan "puhua" (johon voidaan viitata muualta)
2. Käytä HTTP URI:ja niin, että ihmiset voivat etsiä ja tarkistaa niitä nimiä	2. tarjota käsitteeseen liittyvää tietoa sieltä, mistä sitä luontevasti etsisi
3. Kun joku tarkistaa antamasi URI:n, tarjoa hyödyllistä informaatiota käyttämällä standardeja (RDF, SPARQL)	3. tehdä helpoksi lisätiedon löytämisen nimetyistä objekteista ja resursseista
4. Sisällytä linkkejä muihin URI:hin, jotta he voivat löytää lisää asioita	4. muodostaa käsitteiden välisiä suhteita – luoda erillisten tiedonpalasten sijaan koko ajan laajeneva verkko

Riippumatta siitä, missä formaatissa tieto julkaistaan ja missä formaatissa se alunperin on, RDF on käyttökelpoinen malli yhdistämään tietovarantoja keskenään internetin välityksellä. Teknologiana RDF mahdollistaa asioiden ja konseptien helpon linkittämisen toisiinsa sekä toisistaan riippumattomien ja erillään suunniteltujen järjestelmien myöhemmän yhdistämisen toisiinsa. Koska RDF mahdollistaa saman datan kuvaamisen erilaisilla sanastoilla, voidaan yhteentömmivuutta harmonisoinnin ja sanastotyön avulla lisätä siellä, missä se kannattaa kustannustehokkaasti tehdä.

Avoimen datan julkaisun yhteydessä ei ole tarpeen etsiä yhteisiä standardeja ja tapoja kuvailla kaikkea dataa, kuten vaikkapa koulua koskevaa dataa. Yksilöillä ja organisaatioilla on hyvin poikkeavia näkökulmia. Kaupungin tilapalvelun rekistereistä koulu löytynee vuokralaisena, opetusministeriöllä on paljon kouluihin liittyvää dataa ja yksittäisellä koululla itsellään on dataa muun muassa aukioloajoista. Ei ole syytä olettaa, että kaikki toimijat alkaisivat käyttää yhtä standardoitua sanastoa kuvaamaan koulua. Jos tällainen sanasto saataisiinkin sovittua, ei se kompromissina välttämättä enää palvelisi kunnolla mitään tahoa.

Yleisesti hyväksytyt kuvailusanastot lisäävät merkittävästi datavarantojen yhteiskäyttöisyyttä ja siksi niiden vapaaehtoista syntymistä kannattaa tukea. Internetissä toimiva pienimmän yhteisen nimittäjän voima (kohta 6.3) ei ole millään tavalla estänyt yksilöitä ja yhteisöjä rakentamasta yhteisen pohjan päälle heitä tukevia yksityiskohtaisempia käytäntöjä. Esimerkiksi www-sivujen merkintäkieli HTML on hyvin laajasti hyväksytty standardi, mutta sen päälle on vapaaehtoisesti ja tarvelähtöisesti kehitetty suppeampia käytäntöjä, kuten mikroformaattit, joilla vaikkapa osoitetiedot voidaan merkitä HTML:n sekaan koneymmärrettävässä muodossa. Jos tällaisia suppeampia standardeja käytetään, on tärkeää toimia niin, että ne eivät millään tavalla rajoita muun sivun käyttämistä. Sivusto näkyy, vaikka selain ei mikroformaattia ymmärtäisikään eikä yhden mikroformaatin käyttäminen yhdessä kohtaa sivustoa edellytä, että koko sivustolla käytettäisiin kaikkia mikroformaatteja.

Sanastotyössä RDF tarjoaa tasapainoa helppouden ja standardoinnin hyötyjen välillä. Dataa julkaisevilla yksilöillä tai organisaatioilla on täysi vapaus valita, miten ne kuvailevat informaatiotaan, mutta samalla niillä on mahdollisuus jakaa omia sanastojaan ja uudelleenkäyttää osia toisten luomista sanastoista sekä luoda kuvailutietoon sopiva yhdistelmä yleisiä standardeja ja omia näkemyskäsityksiä. Toinen RDF-julkaisua puoltava tekijä on se, että siihen on jälkikäteen helppo yhdistää sanastoja.

Linkitetyn datan verkko on suunniteltu kasvamaan ja kehittymään orgaanisesti samalla tavalla kuin linkitettyjen dokumenttien verkko on kasvanut. Se kasvaa, kun ihmiset ja organisaatiot toisiltaan kysymättä spontaanisti lisäävät omia resurssejaan verkkoon ja linkittävät niitä toisiinsa. Se kasvaa, kun toimijat luovat omia sanastojaan kuvailemaan asioita. Kehitykseen kuuluu itseorganisoituvuus ja sekavuus; linkkejä rikkoutuu, uusia syntyy, sanastoja yhdistyy ja jakautuu.

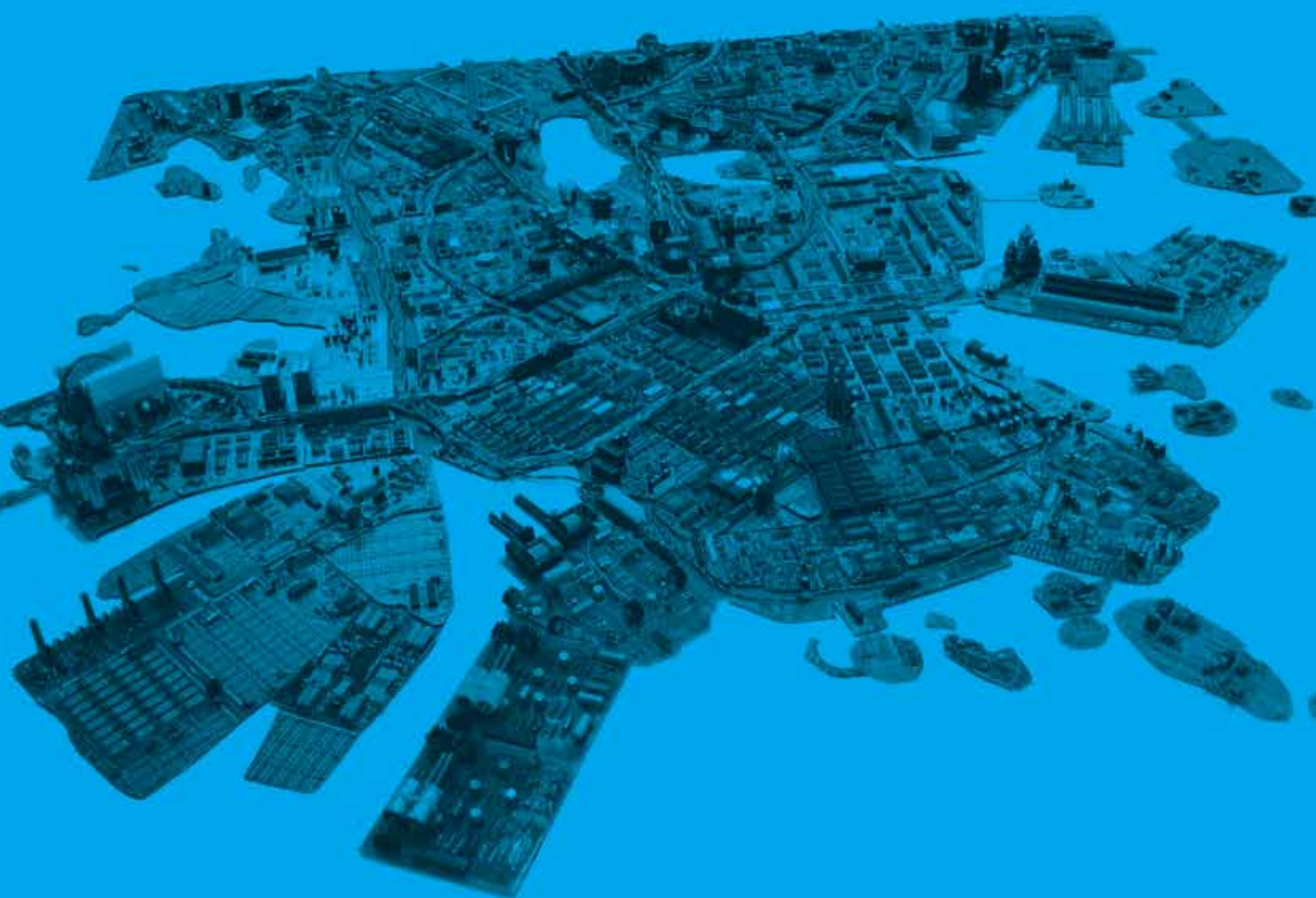
Case: Suomalaista semanttisen webin osaamista – TerveSuomi ja Kirjasampo

Semanttisessa webissä kuvaillaan tiedon keskinäisiä suhteita tietokoneen ymmärtämässä muodossa, jotta www-sisältö olisi helpommin koneellisesti käsiteltävää. Semanttinen web oli kaikkien huulilla muutama vuosi sitten. Vähään aikaan siitä ei ole puhuttu niin paljon, mutta se on edelleen internetin tutkituimpia ja nopeimmin kehittyviä alueita. Sen käytännön merkitys kaiken kohun jälkeen on kasvamassa vauhdilla ja nyt keskitytään toimiviin sovelluksiin. Kansainvälisesti Suomi on yksi semanttisen webin edelläkävijämaita. Tietovarantojen julkaisemiseen liittyen keskeistä semanttisen webin ajattelutavassa on linkitetty data.

TerveSuomi.fi on julkishallinnon ja järjestöjen yhteinen terveystiedon portaal, joka perustuu semanttisen webin tekniikoille ja mukana olevien toimijoiden tarjoamaan avoimeen dataan. Kun sisällöt varustetaan yhtenäisillä metatiedoilla, koneet osaavat yhdistää toisiinsa kuuluvia asioita yhteen. Näin palvelua käyttävä saa tarkempaa täsmätietoa sekalaisten hakutulosten sijaan.

Lähiaikoina valmistuva Kirjasampo-kaunokirjallisuussivusto on ensimmäinen valtakunnallinen kirjastoverkkopalvelu, joka hyödyntää semanttista webiä. Kirjasampo sisältää metadatan kaunokirjallisuudesta (myös vanhasta ja aiemmin vähälle sisällönkuvailulle jääneestä kirjallisuudesta). Rajapinnoiltaan siitä on tarkoitus tulla avoin.

*Eikö olisi hienoa, että kaikki
julkishallinnon hallussa oleva
tieto, mikä on julkista, olisi
myös helposti löydettävissä ja
saatavissa yhdestä paikasta?*



6. Avoimen datan infrastruktuuri

Tässä luvussa esittelemme idean datakatalogeista sekä datan avaamisessa avustavista poikkihallinnollisista toimijoista, joita kutsutaan tässä yhteydessä clearing house -organisaatioiksi.

Poikkihallinnolliset toimijat ovat datanavauksen konsultteja, jotka auttavat yhdistämään osapuolten intressejä. Datakatalogi konkreettisen verkkosivustona on tällaisen poikkihallinnollisen tukioorganisaation sekä datan julkaisijoille että uudelleenkäyttäjille tarjoama palvelu. Poikkihallinnollinen toimija voi olla kansallisella tasolla oma tietohallintovirastonsa, kuten Iso-Britannian Office of Public Sector Information (OPSI), tai seudullisesti tai kaupungin sisällä toimiva hallintokuntien rajat ylittävä tukioorganisaatio.

Datakatalogi on strukturoitu metadatarekisteri, johon on yhdistetty metadatta useampien julkisten organisaatioiden hallussa olevista aineistoista. Datakatalogit voivat olla kansallisia (mm. suomi.fi/datakatalogi ja data.gov.uk), seudullisia (Washington D.C. tai mahdollinen Helsinki Region Infoshare), kaupunkien ylläpitämiä (San Francisco) tai yksityisten tahojen ylläpitämiä (Sunlight Foundation - National datacatalog). Ideaalitapauksessa datakatalogin ylläpitäjä voi tarjota tukea datan julkaisijoille esimerkiksi liittyen lisensointiin tai ohjelmallisesti käsiteltäviin rajapintoihin ja formaatteihin. Datakatalogia ylläpitävä taho voi myös toimia kontaktina datan hyödyntäjiin organisoimalla esimerkiksi uudelleenkäyttöön kannustavia kilpailuja tai keräämällä kokemuksia, palautetta ja datan avauspyyntöjä.

Case: Kevyesti alkuun – Suomen ensimmäinen datakatalogi

Suomi.fi:n toiminta vuonna 2009 järjestetyn Suomen ensimmäisen Apps for Democracy -kilpailun yhteydessä on esimerkki nopeasta reagoinnista ja toimintaan ryhtymisestä. Kilpailussa kehitettiin palveluita julkisten koneluettavien tietolähteiden varaan. Tarvittiin paikka, johon koota linkit aineistoihin. Suomi.fi:n toimitus otti tehtäväkseen ylläpitää listaa tietolähteistä aivan tavallisella verkkosivulla Laatusivustoon -projektin sivuston yhteydessä. Toteutus ei ollut tietolähteiden esillepanon näytävyyden tai löydettävyyden kannalta optimaalinen, mutta se saatiin aikaiseksi lyhyellä varoitusaajalla.

Datakatalogin esiaste (<http://suomi.fi/datakatalogi>) vaikutti enemmän kuin Suomi.fi:n toimittajat ehkä arvaavatkaan. Kilpailu ja katalogi osoittautuivat tärkeiksi välineiksi datan avaamisen konkretisoimisessa. Nopea, kevyt ja hyvin ajoitettu reagoiminen tuotti ehkä suuremman tuloksen, kuin laajempi mutta hitaampi toiminta.

Sivusto ei ollut pelkästään lista tietolähteistä, vaan se oli signaali siitä, että hallinto tukee kilpailua ja haluaa avata datavarantojaan. Suomi.fi:n toimitus toimi välittäjänä, joka puhui sekä muiden hallinnon organisaatioiden että kilpailun järjestäjien kanssa. Ilman hallinnon tukea kilpailu olisi saanut erilaisen vireen. Sivuston julkistaminen on hyvä esimerkki siitä, minkälaisia kerrannaisvaikutuksia pienellä teolla voi olla avoimen datan ekosysteemin kehittymisen kannalta Suomessa.

6.1 Datakatalogi – kaikki julkinen data yhdeltä luukulta

Eikö olisi hienoa, että kaikki julkishallinnon hallussa oleva tieto, mikä on julkista, olisi myös helposti löydettävissä ja saatavissa yhdestä paikasta?

Katalogien avulla datavarantojen olemassaolo tehdään tunnetuksi potentiaalisille datan hyödyntäjille. Tarve julkisten datavarantojen löydettävyyden parantamiseksi on huomattu myös EU:n PSI-direktiiviä (jakso 3.5.1) seuranneen työryhmän suosituksissa: «Käytännöllisiä hankkeita datavarantojen rekistereiden ja muun PSI-infrastruktuurin luomiseksi tulisi tukea kansallisesti ja myös rajoja ylittäen Euroopan tasolla». Suositustekstin mukaan datakatalogit, jotka sisältävät rakenteellisessa muodossa metatietoa sekä julkaistuista että toistaiseksi vielä julkaisemattomista hallinnon datavarannoista, ovat välttämättömiä hallinnon tietovarantojen uudelleenkäytön lisäämisessä, ja ne tuottavat samalla merkittäviä hyötyjä dataa hallinnoiville organisaatioille itselleen.

Vuoden 2009 aikana tämä ajatus yhden luukun periaatteella toimivista julkisen datan katalogeista löi maailmalla läpi kenties Obaman hallinnon esimerkkiä seuraten. Yhdysvaltain hallinnon data.gov julkaistiin toukokuussa 2009 ja sen jälkeen julkaistiin maailmalla kymmeniä julkishallinnon datavarantoja listaavia ja jakelevia datakatalogeja valtioiden, kaupunkien ja kaupunkiseutujen tasolla (esim. San Francisco, Lontoo, New York). Monissa paikoissa, missä hallinto ei vielä ole datakatalogia tehnyt, ovat kansalaiset olleet aktiivisia, kuten Ruotsissa (opengov.se). Vuotta 2009 voidaan hyvällä syyllä kutsua kansainväliseksi julkishallinnon datavarantojen julkaisemisen vuodeksi.

Muutamaa poikkeusta lukuun ottamatta ennen vuotta 2009 julkaistut datakatalogit keskittyivät rajatun esimerkin ainoastaan paikkatietoon, vaativat rekisteröitymistä tai olivat tarjolla rajatuille kohderyhmälle, kuten tutkijoille. Datakatalogien julkaisun lisäksi Yhdysvaltain ja Britannian hallitukset myös sitoutuivat strategisesti avoimen datan tuottamiseen (Digital Britain ja Open government directive).

Case: data.gov.uk

Toistaiseksi ehkä näyttävin hallinnon datakatalogi on 19.1.2010 julkisena betaversiona avattu Ison-Britannian data.gov.uk. Tämä www:n kehittäjän Tim Berners-Leen johdolla kuudessa kuukaudessa koottu sivusto päihittää USA:n data.govin mennessä tullen. Julkaisuhetkellä Britannian katalogissa on noin 3000 tietoa-aineistoa, mikä on kolme kertaa enemmän kuin data.gov:ssa.

Datan määrän ohella Iso-Britannian katalogin sisältöä voidaan pitää myös kiinnostavampana. USA:n data.gov:ia on sen julkaisusta asti kritisoitu siitä, että siellä on vain sellaisia aineistoja, jotka eivät vahingossakaan herätä poliittista keskustelua. Britannian katalogista sen sijaan löytyy mm. armeijassa tapahtuneiden kuolemantapausten tilastoja ja muuta läpinäkyvyyden mielessä kiinnostavaa aineistoa.

Britannian katalogi on myös rakennettu tukemaan koneluettavuutta ja semanttisen webin ominaisuuksia. Suoraan katalogin etusivulta pääsee sivulle, jonka kautta voi selaimessa tehdä tietokantahakuja SPARQL nimisellä kyselykielellä. Myös hakutuloksia – esimerkiksi brittiläisiä kouluja – voi tutkia selaimessa ilman että omalle koneelle tarvitsee ladata mitään. Katalogissa käytetyn ratkaisun merkitys on, että siinä käytettyyn teknologiaan perehtyminen onnistuu erittäin nopeasti.

6.1.1 Datakatalogien tekninen yhteensopivuus

On luontevaa, että valtionhallinnolla on oma virallinen datakataloginsa, jonka rinnalla toimii paikallisia ja aihealuekohtaisia katalogeja. Julkisen datan tuottajina seudulliset ja kuntatason toimijat ovat erittäin merkittävässä asemassa, sillä suuri osa uudelleenkäytön kannalta kiinnostavasta datasta löytyy niiden organisaatioilta. Eri maissa datakatalogien yleistymisen on saanut julkisuutta yleensä korkean profiilin kansallisen datakatalogin myötä. Kehitys on kuitenkin usein käynnistynyt jo aiemmin paikallisen tason datakatalogeilla. Esimerkiksi Kanadassa on jo lukuisia eri kaupunkien omia katalogeja, mutta kansallinen sivusto on siellä vasta kehitteillä.

Virallisten tahojen ohella myös yksityiset tahot ovat perustaneet omia julkisten datavarantojen katalogeja, kuten Sunlight Foundation Yhdysvalloissa ja Ruotsin opengov.se. Näitä katalogeja on perustettu toisaalta sen takia, kun julkista katalogia ei ole ollut ja toisaalta sen takia, että ohjelmoijat ja kansalaiset saattavat vaihtaa ajatuksiaan vapautuneemmin omalla datayhteisösivulla, jonka ylläpito on viranomaisista riippumatonta.

Kansallinen taso ei missään tapauksessa ole julkishallinnon datavarantojen hyödyntämisessä korkein tarkastelutaso. Erityisesti EU:n tasolla on kiinnostusta jäsenmaiden väliseen yhteistyöhön ja sitä kautta monikansallisiin julkishallinnon datakatalogeihin ja ajan myötä myös aineistojen harmonisoimiseen niiden käytettävyyden lisäämiseksi.

Iso-Britannian datakatalogi on toteutettu kokonaan avoimen lähdekoodin ratkaisuilla. Julkaisujärjestelmänä on Drupal ja aineistojen metadatan hallinta on toteutettu Open Knowledge Foundationin vuodesta 2006 lähtien kehittämällä CKAN-alustalla. Tähän mennessä kaupallisista järjestelmätoimittajista ainoastaan Microsoft on julkaissut julkishallinnon datakatalogeja varten oman tuotteensa *Open Government Data Initiative*, joka on otettu ensimmäisenä käyttöön Edmontonin kaupungissa.

Datakatalogien yhteensopivuus liittyy oikeastaan keskeisesti kysymykseen siitä, mitä metadataa julkishallinnon datalähteistä halutaan kerätä. Tähän ei ole vielä olemassa valmiita laajasti hyväksyttyjä standardeja. Niiden valmistumista ei myöskään kannata jäädä odotelemaan, sillä ne muotoutuvat olemassaolevien ja syntyvien katalogien ja käytäntöjen kautta. Tässä vaiheessa riittää, että huolehditaan Suomeen syntyvien katalogien keskinäisestä yhteensopivuudesta ja yhteentoimivuudesta merkittävimmän eurooppalaisen katalogin eli data.gov.uk:n kanssa.

6.2 Taustatyö datakatalogin takana

Päällepäin edellä kuvatut datakatalogit ovat parhaimmillaan yhden luukun periaatteella toimivia sivustoja, joiden kautta pääsee kiinni julkiseen tietoon. Pintaa syvemmältä datakatalogit edustavat ns. clearing house -ajattelua. Clearing house -organisaatio on datan käyttäjien ja tuottajien välissä toimiva taho, jonka tehtävänä on kerätä, säilyttää ja levittää informaatiota, metadataa ja dataa. Clearing house on finanssitoiminnasta yleistynyt termi välittäjäorganisaatiolle. Suomeksi se voisi olla vaikka selvitystoimisto, mutta käytämme tässä oppaassa englanninkielistä termiä.

Datan hyödyntäjille ja jatkojalostajille tällainen clearing house -organisaatio saattaisi näyttäytyä hyvin ylläpidettynä katalogina julkishallinnon tietovarantoihin. Yhdestä paikasta löytyisi kaikki informaatio, mikä lain mukaan on julkista. Datakatalogin taustalla toimiva clearing house -organisaatio selvittäisi työkseen lukuisten muiden organisaatioiden kanssa teknisiä, juridisia ja organisatorisia kiemuroita sekä harmonisoisi ja kuvailisi yhdenmukaisesti eri lähteistä tulevaa

dataa. Se voisi myös hoitaa julkisen tietokannan ylläpidon ja kehittämisen, niin että lukuisat organisaatiot, jotka nykyisin ylläpitävät ns. operatiivisia tietokantoja omia tarkoituksiaan varten voisivat ulkoistaa tämän toiminnon clearing house -organisaatiolle ja keskittyä informaation jalostamiseen.

Clearing house olisi siis vastuussa siitä, että raakadata olisi koneluettavassa muodossa saatavissa kaikille sitä tarvitseville ja siitä kiinnostuneille niin hallinnon sisällä kuin ulkopuolellakin. Sillä ei olisi mitään velvoitteita jalostaa dataa eikä tuottaa ns. perusnäkymiä, joissa data on esitetty luettavassa muodossa. Nykyisin perusnäkökymien tekeminen on viranomaisen velvollisuus ja koneluettavien rajapintojen tekeminen jää usein lapsipuolen asemaan. Raakadatan tarjoaminen rajapintoihin voitaisiin ulkoistaa clearing house -organisaatiolle ja viraston tehtäväksi jäisi edelleen vain se, mikä sillä on nykyäänkin: peruspalveluiden tuottaminen ja informaation jalostaminen.

Toki informaation jalostuksen myötä syntyy sellaista uutta tietoa, jota ei voida enää kutsua raakadatakksi, mutta joka voidaan haluttaessa edelleen tarjota rajapintojen kautta maksutta eteenpäin. Tällöinkin virasto voisi ulkoistaa tämän jalostetun informaation tarjoamisen clearing house'lle.

6.2.1 Julkisen datan avaamispalvelu

Vaikka julkinen data on periaatteessa saatavilla, saattaa sen käytännön hyödynnettävyydessä ilmetä ongelmia, kun dataa halutaan kopioida, yhdistää muihin datalähteisiin, jalostaa tai julkaista uudelleen. Haasteet saattavat liittyä esimerkiksi maksullisuuteen, käyttöehtoihin tai formaatteihin, jossa julkinen informaatio on tarjolla (2.2 avoimen datan mittareita).

Datan hyödyntäjien kannalta olisi luontevaa voida ilmoittaa käytännön ongelmista keskitetysti yhdelle taholle, jolla olisi valtuuksia ja keinoja pyrkiä yhdessä datan tarjoajan kanssa parantamaan aineiston käytettävyyttä. Iso-Britanniassa edellä mainittu OPSI tarjoaa julkisen datan avaamispalvelua (Public Sector Information Unlocking Service), jonne voi ilmoittaa mm. aineistoista, jotka puuttuvat data.gov.uk katalogista tai ovat muutoin hankalasti käytettyjä. Julkisen datan avaamispalvelua voisi verrata esimerkiksi kuluttajavalituslautakuntaan, mutta ainakin Britanniassa se on brändätty positiivisempaan sävyyn. Palveluun jätetyt ilmoitukset ovat kaikille avoimia ja muut käyttäjät voivat kommentoida ja äänestää ehdotuksia ja niitä ratkotaan tärkeysjärjestyksessä.

6.3 Suomalaisia hankkeita

Julkinen datan infrastruktuurin kehittäminen on Suomessa juuri nyt erittäin ajankohtaista. Syksyllä 2010 valmistuvan kansallisen tietoyhteiskuntastrategian valmistelussa yhtenä kuudesta teemasta on esitetty «Tiedon roolia yhteiskunnan perusinfrastruktuurina on tuettava» (Otakantaa.fi 2010). Työ- ja elinkeinoministeriön tuoreessa 16.2.2010 julkaistussa käyttäjälähtöisen innovaatiopolitiikan toimenpideohjelmassa (TEM 2010) esitetään toimenpiteksi: «Arvioidaan sidosryhmien kanssa tarve perustaa kansallinen julkisen sektorin tiedon hyödyntämistä koskeva yhteyspiste, jonka tehtävänä olisi mm. jakaa tietoa, tarjota opastusta sekä verkottaa toimijoita.»

Kuten luvun alussa kerroimme, vuonna 2009 Suomi.fi otti ansiokkaasti pallon haltuunsa julkishallinnon datakatalogin kehittämisessä. Kevättalvella 2010 suomi.fi kehittää seuraavaa versiota datakatalogista. Tavoitteena on rakentaa virallinen paikka, mihin hallinnon edustajat voisivat ilmoittaa omia datavarantojaan. Valtion hallinnoiman katalogin ohella kehitetään parhaillaan myös kirjastot.fi:n tuella yhteisöllistä pääasiassa ohjelmoijille ja palveluideoijille suunnattua

opengov.fi -sivustoa. Siellä datan hyödyntäjät voivat keskustella ja jakaa epä-määräistäkin tietoa tyyliin, «*Helsingilläkin on varmaan olemassa rekisteri kaupungin istutetuista puista*». Tavoitteena on kevyt sivusto, joka toimii yhteen muiden virallisempien katalogien kanssa. Pääkaupunkiseudun kuntien kesken on käynnistymässä seututiedon harmonisointiin ja maksuttomaan jakamiseen tähtäävä «Helsinki Region Infoshare» -hanke.

Vaikka yleiset julkisen datan katalogit ovat verrattain uusi ilmiö, on esimerkiksi paikkatiedoille ollut hakupalveluita käytössä jo useita vuosia. Maanmittauslaitos on parhaillaan ottamassa käyttöön uuden sukupolven *paikkatietohakemisto.fi*-palvelua, joka perustuu avoimen lähdekoodin GeoNetwork -ohjelmistoon. Paikkatietojen osalta INSPIRE-direktiivi on määritellyt paikkatietojen metadatan yhdenmukaisesti (GEMET-sanasto) ja kuvannut niitä varten kehitettävät kansalliset ja euroopanlaajuisen hakupalvelun. Direktiivin toimeenpanon yhteydessä on huomioitu organisatorisia ja hallinnollisia asioita, jotka eivät ole vain paikkatietosidonnaisia. Niitä voidaan soveltaa avoimen datan infrastruktuurin kehittämisessä.

Metadatan ja sanastojen määrittelyssä edellisessä luvussa esitetty Linked-data-lähestymistapa on eduksi. Julkaisemalla hallinnon datakatalogin RDF-muodossa kaikkea ei tarvitse määritellä etukäteen, vaan tarvittaessa uusia määrittelyitä voidaan ottaa käyttöön vaikka käsite kerrallaan. Aalto yliopiston semanttisen webin tutkimusryhmä kehittää parhaillaan semanttisen datakatalogin prototyyppiä. Myös käynnissä olevassa valtiollisen tason tietoarkkitehtuuriprojektissa (VALTASA) noudatetaan semanttisen webin ajattelutapaa (VM 2009).

*Suomella on juuri nyt ainutlaatuinen
mahdollisuus kohdentaa
voimavaroja avoimeen dataan.
Maassa valmistellaan uutta
tietoyhteiskuntastrategiaa ja pian
myös uutta hallitusohjelmaa.*



Tämä opas on kirjoitettu yksittäisen julkishallinnon organisaation näkökulmasta ja tarkoitettu tukemaan koneluettavan tiedon, eli datan, hyödyntämisen lisäämistä Suomessa sekä hallinnon organisaatioiden välillä sekä laajemmin yhteiskunnassa. Datavarantojen avaamisessa ja tarjoamisessa maksutta jatkojäljestettäväksi on paljon tehtävissä nykyisen lainsäädännön ja toimintalinjausten puitteissa. Oppaan kirjoittamista edeltäneessä kansainvälisessä vertailussa ja julkishallinnon haastattelukierroksella nousi esiin datan hyödyntämiseen liittyviä tarpeita ja toimintamalleja, jotka on hyvä huomioida kansallisella tasolla. Ne tuodaan tässä luvussa kootusti esiin toimintaehdotuksina. Esimerkiksi huomio datan tulkintaan liittyvästä lukutaidosta ja ajatus sen sisällyttämisestä koulutukseen nousi toistuvasti esiin haastatteluissa.

Avointa dataa tuottavan ja hyödyntävän ekosysteemin synnyttäminen vaatii monenlaista ja -tasoista tukea: rakenteellisia ja lainsäädännöllisiä muutoksia, toimintatapojen uudistamista ja tiivistä vuorovaikutusta eri tahojen välillä. Tarvitaan myös innokkaita ihmisiä ja organisaatioita, jotka ovat valmiita kokeilemaan ja oppimaan kokeiluista. Datan avoimuutta on hedelmällistä tarkastella koko yhteiskunnan laajuudessa ja kytkeä se kansallisiin tavoitteisiin demokratian, innovaatiotoiminnan ja hallinnon tehokkuuden lisäämiseksi:

- 1)** Läpinäkyvyyden ja demokratian osalta avoin maksuton data linkittyy vuonna 2009 käynnistyneen Sähköisen asiointin ja demokratian vauhdittamisohjelman (SADe) tavoitteisiin kansalaisten osallistumismahdollisuuksien parantamisesta (VM 2009b).
- 2)** Liiketoiminnan ja innovaatioiden osalta avoin maksuton data tukee sitä hyödyntävän edelläkävijämarkkinan synnyttämistä (TEM 2008) ja erityisesti vuonna 2010 julkaistun kysyntä- ja käyttäjälähtöisen innovaatiopolitiikan toimenpideohjelman tavoitteita (TEM 2010).
- 3)** Hallinnon sisäisen tehokkuuden osalta avoin maksuton data tukee Valtiotason arkkitehtuurit -hankkeen linjausta tiedosta yhteiskäyttöisessä pääomana (VM 2009c).

Osa tämän luvun toimenpide-ehdotuksista, kuten maksuperustelain tarkastelu ja kansallisten datakatalogien perustaminen, on mainittu myös muualla ja työjoidenkin toteuttamiseksi on jo käynnistynyt. Mukana on myös joukko uusia ehdotuksia. Ehdotukset on ryhmitelty (1-4) avoimen datan perusinfrastruktuuria rakentaviin, (5-7) lainsäädännön muutosta ja ohjausta koskeviin, (8-12) hallinnon sisäisiä käytäntöjä tukeviin, (13-15) ja avoimesta datasta saatavien hyötyjen edistämiseen, sekä (16-18) Suomen kansainvälistä edelläkävijyyttä ja tahtotilaa mahdollistaviin. Loppuun on koottu joukko esimerkkejä, jotka konkretisoivat ehdotuksia.

Kaikkien allamainittujen toimenpide-ehdotusten toteutuminen ilman erityistä tärkeysjärjestystä tukisi merkittävästi suomalaisen avoimen datan ekosysteemin kehittymistä. Eri toimijoiden on varmasti helppo mainittujen toimenpide-ehdotusten toteutuminen ilman erityistä tärkeysjärjestystä tukisi merkittävästi suomalaisen avoimen datan ekosysteemin kehittymistä. Eri toimijat voivat tarttua alla olevista ehdotuksista heitä lähinnä oleviin ja edistää niitä.

7.1 Avoimen datan perusinfrastruktuurin synnyttäminen

Oppaan ehdottamassa tavoitetilassa harmonisoitu ja organisoitu data on helposti saavutettavissa, ymmärrettävästi dokumentoitu, kommentoitavissa ja laajasti käytössä. Datan käyttöönottoon ja mahdolliseen maksamiseen liittyvät prosessit on automatisoitu, eikä datan uudelleenkäyttöön liity esteitä. Tämä edellyttää, että datavarantojen jakeluun on olemassa kattava tekninen, organisatorinen ja juridinen infrastruktuuri.

Koska raakadataa syntyy lähes kaikissa organisaatioissa, pitäisi avoimen datan julkaisemiseen olla resursseja ja osaamista kaikissa julkishallinnon organisaatioissa. Kun avoimen datan perustukset ja osaaminen on kertaalleen rakennettu, ei niiden ylläpitäminen jatkossa enää vaadi yhtä paljon resursseja kuin selvittelyvaiheessa.

Suomessa on jo olemassa useita virallisia sekä yhteisöjen ylläpitämiä datakatalogeja^[1]. Olennaista on näiden katalogien yhteentoimivuus tekniikan, käyttöehtojen ja ylläpitokäytäntöjen osalta. Tavoitteeseen päästään katalogeja ylläpitävien henkilöiden yhteistyöllä. Jotta datan saatavuutta ja käyttöä tukeva ekosysteemi syntyisi, Suomeen tarvitaan lisäksi data-aineistojen uudelleen käytön edistämistä ja julkaisemisen valvonnasta vastaava taho sekä tietovarantojen kuvailun yhteiset periaatteet.

1. *Clearing house -organisaatiot edistämään datan uudelleenkäyttöä ja jakelua*

Sekä kansallisella että suurimpien kaupunkiseutujen^[2] tasolla tulisi julkisen datan jakelu ja uudelleenkäytön edistäminen vastuuttaa poikkihallinnolliselle toimijalle (clearing house; jakso 8.2). Toimijalla olisi riittävät resurssit ja osaaminen auttaa muita viranomaisia, yrityksiä ja kansalaisia niin datavarantojen julkaisussa kuin niiden uudelleenkäytössäkin.

2. *Datakatalogit edistämään aineistojen löydettävyyttä*

Kehitetään valtion ja kaupunkiseutujen tasolla julkisten datavarantojen katalogeja (jakso 6.1). Varmistetaan uusien katalogien yhteentoimivuus ja aineistojen linkittyminen (jakso 5.6) kehitteillä oleviin kotimaisiin^[1], sekä kansainvälisiin katalogeihin. Kansallisen tason datakatalogiin panostetaan ja siitä pyritään kehittämään lippulaiva -katalogi, joka levittää tietoa avoimesta datasta ja haastaa positiivisesti maailman muita katalogeja, esimerkiksi olemalla niitä käytettävämpi ja yhteisöllisempi.

3. *Suomalaisten tietovarantojen kuvailu ja määrittely*

Määritellään ja julkaistaan tiedot suomalaisten julkisten datavarantojen sisällöstä, hallinnosta ja kehittämisvastuista^[3]. Laaditaan nopeasti periaatteet, joilla tietovarantojen kuvailu voidaan tehdä kootusti linked data -paradigman mukaisesti (jakso 5.6).

^[1] Suomalaisia datalähteitä kokoavia verkkopalveluita ovat ainakin: www.paikkatietoikkuna.fi, www.suomi.fi/datakatalogi ja www.opengov.fi

^[2] Helsingin seudulla käynnisteillä oleva Helsinki Region Infoshare -hanke on esimerkki tietovarantoihin liittyvästä seudullisesta yhteistyöstä.

^[3] Laki tie- ja katuverkon tietojärjestelmästä 28.11.2003/991 on esimerkki laista, joka määrittelee data-aineiston ylläpitämiseen liittyvät vastuut.

4. Julkisdata-valvontaelimen perustaminen

Määrätään viranomaisen valvomaan julkisuuslain mukaisten aineistojen julkaisemista avoimen datan periaatteiden mukaisesti (jakso 2.2). Viranomaisen myös auttaa aineistojen julkaisemiseen liittyvässä lisensoinnissa ja juridisissa kysymyksissä. Viranomaiselle voidaan ilmoittaa julkisten aineistojen datan avoimuuteen liittyvistä, mahdollisista ongelmatilanteista^[4].

7.2 Lainsäädännön muutos ja kansallinen ohjaus

Suomalaisten tietovarantojen hyödynnettävyyden huomattavan parantamisen näkökulmasta on tärkeää selvittää mahdollisuudet avoimen datan maksuttomaan tarjoamiseen. Maksuperustelain ja -asetuksen uudistamistarpeen arvioimista on viimeksi ehdottanut tilastotoimen tehostamista ja alueellistamista valmistellut työryhmä (VM 2010). Mahdolliset tiedon hinnoittelun muutokset kaipaisivat tuekseen yhteiskunnallista keskustelua ja taloudellisten vaikutusten arvioimista. Mahdollinen maksutulojen väheneminen edellyttäisi oppaan näkemysten mukaan suoran budjettirahoituksen kasvattamista.

Tietovarantojen hyödynnettävyys tulisi lainsäädännön lisäksi huomioida organisaatioiden toimintaohjeissa. Datan avaamisen tueksi tarvittaisiin ohjeistusta, josta yksittäinen hallinnon organisaatio niin valtio- kuin kuntatasolla pystyisi hahmottamaan sekä datavarantojensa luonteen että soveliaimman toimintamallin niiden suhteen. Tällaisen ohjeistuksen laatiminen voitaisiin tehdä esimerkiksi eri ministeriöiden ja isoimpien kuntien toimesta.

Avoimeen dataan liittyy myös yhteiskunnallisen keskustelun tarve yksityisyyden suojan harmaista alueista. Kaiken sen datan, mikä nykyisen lainsäädännön puitteissa on julkista, ei kenties tulisi enää olla julkista, jos se on helposti saatavissa koneluettavassa muodossa.

5. Maksuperustelain uudistaminen

Valtion maksuperustelakia tulisi julkisten digitaalisten aineistojen osalta muuttaa niin että ne ovat lähtökohtaisesti maksuttomia sekä hallinnon sisällä, että hallinnon ulkopuolisille toimijoille.

6. Esiselvitys datan maksuttomuuden vaikutuksista

Selvitetään datan maksuttomuuden vaikutuksia ja erilaisten budjetointimallien kokonaisedullisuutta. Ehdotamme talouslähtöistä esiselvitystä 1) nykymallin kuluista ja tuotoista sekä 2) maksuttomuus -vaihtoehdon arviointia. Selvityksessä datan tarjoamisen vaikutuksia tulee tarkastella sekä yksittäisten virastojen tasolla, että koko valtion mittakaavassa (jakso 4.3.1).

7. Datan avaamisen kirjaaminen tulosohtaukseen

Datavarantojen avaaminen ja kehittäminen tulisi ottaa mukaan hallinnon organisaatioiden tulosohtaukseen. Ensivaiheessa tavoitteeksi voidaan esimerkiksi ottaa omien datavarantojen inventoiminen. Oppaassa esitettyä etenemismallia (jakso 2.3) voidaan käyttää pohjana tavoitteiden määrittelyssä.

^[4] Britanniassa toimii kansallinen Office of Public Sector Information, joka ylläpitää Unlocking-palvelua: <http://www.opsi.gov.uk/unlocking-service/>

7.3 Hallinnon sisäiset käytännöt tukemaan avointa dataa

Oikeusministeriössä valmistellaan sosiaalisen median opastusta virkamiehille. Avoin data ja sosiaalinen media tukevat toisiaan ja lisäävät yhdessä hallinnon organisaatioiden osaamisen ja tietovarantojen vaikuttavuutta. Olisi varsin luontevaa sammalla kertaa tarkastella avoimen datan ja sosiaalisen median sisällyttämistä hallinnon organisaatioiden toimintaan ja strategiaan^[5].

8. Hallinnon asiakirjojen ja aineistojen lisensointikäytäntöjen määrittäminen

Määritellään julkisuuslain määrittelemille viranomaisten asiakirjoille kansainvälisten internet-käytäntöjen kanssa yhteensopivat ja uudelleenkäyttöön kannustavat käyttöehdot.

9. Avoin data ja sosiaalinen media mukaan organisaatioiden strategioihin

Avoimen datan käytänteet ja sosiaalinen media tarjoavat yhdessä kanavan organisaation toiminnasta ja sen tuottamasta tiedosta kiinnostuneiden ihmisten kohtaamiseen. Ne voivat oikein hyödynnettyinä lisätä organisaation vaikuttavuutta ja nopeuttaa oppimista. Niiden avulla organisaatio voi palvella ihmisiä paremmin ja löytää uusia ratkaisuja omiin tarpeisiinsa. Siten avoimen datan ja sosiaalisen median hyödyntäminen tulee sisällyttää organisaatioiden pitkän tähtäimen toimintavalintoihin.

10. Avoimen lähdekoodin huomioiminen hankinnoissa

Avoimen lähdekoodin ohjelmistot ja avoin data tukevat toisiaan. Datan hyödyntäminen helpottuu, jos myös sen käsittelyyn kehitettyjä ohjelmistoja, ohjelmistokirjastoja ja -koodia voidaan jakaa vapaasti^[6]. Avoimen lähdekoodin ohjelmistot suosivat avoimia datan esitysmuotoja.

11. Avoin data mukaan tietojärjestelmien suunnitteluun

Tietojärjestelmiä hankittaessa pitäisi jo hankkeen suunnittelu- ja hankintavaiheessa huomioida datan avoimuus, sekä järjestelmien tuottaman datan, että niiden vastaanottaman datan osalta. Huomioimalla avoin data pitkäjänteisesti, voidaan datan avoimuutta edistää järjestelmien uusimisen yhteydessä.

12. Palveluiden luonnostelun lisääminen hankinnoissa

Datan avaaminen ja avoimen lähdekoodin hyödyntäminen helpottavat organisaation omaan tietojärjestelmään liittyvien kokeilujen tekemistä. Jos organisaatiolla on lisäksi ohjelmointiosaamista käytettävissään, voidaan uusia ratkaisuja ja teknologioita tutkia itsenäisesti ennen hankintaprosessin käynnistämistä. Avoimet ratkaisut mahdollistavat myös prosessin avaamisen avoimen datan yhteisölle. Innovatiivisia ja esikaupallisia julkisia hankintoja (EU 2007b) tulisi voida hyödyntää tässä.

^[5] Suomalaisista organisaatioista YLE on aktiivisesti tuonut sosiaalisen median ja aineistojensa jakamisen mukaan strategiaansa.

^[6] Suomessa on vuoden 2009 alkupuolelta lähtien ollut voimassa JHS-suositus 169 avoimen lähdekoodin ohjelmien käytöstä julkishallinnossa.

7.4 Julkishallinnon avoimen datan hyötyjen jalkauttaminen

Kuten oppaassa on todettu, ei datan avaaminen ole itsearvoista toimintaa. Siitä on hyötyä vasta kun dataa hyödynnetään. Julkishallinnon organisaatioiden toiminnan suuntaamisen lisäksi tulisi eri tavoin pyrkiä tuomaan avoimen datan mahdollisuuksia yleiseen tietoisuuteen ja energisoida avoimen datan ekosysteemin toimijoita. On myös oletettavissa, että dataa hyödyntävien palveluiden ja visualisointien, kuten Veropuu, yleistyessä tarve data-lukutaidon kehittämislle kasvaa.

13. Datan hyödyntämisen esituotantovaiheen rahoitusta

Avoimen datan mahdollisuuksien ymmärtämisen kannalta toimivat palveluesimerkit ovat erityisasemassa. *Apps for Democracy Finland* -kilpailu on tuonut suomalaisia esimerkkejä julkisuuteen. Palveludemojen kehittämiseksi tarvitaan myös hakuprosessiltaan ja raportoinniltan kevyttä demo-rahoitusta^[7].

14. Näkyvät keihäänkärkihankkeet

Käynnistetään välittömästi 3–5 korkean profiilin keihäänkärkihanketta organisaatioiden kanssa, joilla on jo hyvät valmiudet osallistua avoimen datan ekosysteemiin. Hankkeiden yhteydessä avataan laajasti käytettäviä aineistoja yhteiskunnan eri osaalueille, kuten demokraattinen päätöksenteko, kaupunkitieto tai terveys.

15. Avoimen datan tuominen koulutukseen ja yleissivistykseen

Avointa dataa tulisi hyödyntää opetuksessa harjoitusmateriaalina, ajankohtaistamassa ilmiöiden nykytilaa ja konkretisoimassa asioiden mittasuhteita. Datan lähdekritiikkiä ja data-lukutaitoa voitaisiin opettaa esimerkiksi mediakasvatuksen yhteydessä. Järjestetään dataan liittyvää täydennyskoulutusta datan kanssa tekemisissä oleville ammattiryhmille, kuten journalisteille.

7.5 Suomen kansainvälisen edelläkävijyyden mahdollistaminen

Datan avaamisella saavutetaan kansallisesti suurempia etuja, jos Suomi pysyy saavuttamaan edelläkävijämaan aseman. Se tukee yritysten mahdollisuuksia kehittää uusia toimintamalleja ja tuotteita, joita voidaan myöhemmin viedä muihin maihin. Edelläkävijäisyys lisää suomalaisten asiantuntijoiden painoarvoa kansainvälistä avoimen datan infrastruktuuria ja standardeja kehitettäessä sekä helpottaa ulkomaalaisten asiantuntijoiden saamista suomeen.

^[7] AVEK:in DigiDemo-avustus on osoittautunut toimivaksi malliksi demo-vaiheen projektien tukemiseen:
http://www.kopiosto.fi/avek/digidemo/digidemo/fi_FI/yleista/

16. Avoimen datan tutkimus

Panostetaan linkittyneiden tietovarantojen ("big data") tutkimusyksikköön, jossa kehitetään osaamista ja työkaluja datavarantojen laajamittaiseen hyödyntämiseen. Avoin data tulee lisäksi sisällyttää osaksi poikkitieteellisiä tutkimus- ja opetusohjelmia eri näkökulmista: kuten käytettävyys, visualisointi ja yhteiskunnallinen tutkimus.

17. Aineiston yhteistuotannon pilotoiminen

On oletettavissa, että avoimien data-aineistojen yleistyessä kansalaisten kiinnostus aineistojen päivittämiseen ja täydentämiseen lisääntyy. Käynnistetään data-aineistojen yhteistuotantokokeilu, jossa kerätään kansalaisten havaintoja ja mikrokontribuutioita julkishallinnon ylläpitämään aineistoon^[8]. Yhteistuotantomallien kehittämisessä on mahdollista saavuttaa kansainvälisesti edelläkävijän asema, sillä maailmanlaajuisesti tietovarantojen avaamisessa pääpaino on toistaiseksi ollut datan saatavuuden parantamisessa.

18. Suomi viemään datan avoimuutta EU:hun

Suomalaista osaamista ja tavoitteita tuodaan esille osallistumalla EU2020:n lipulaivahankkeeseen Euroopan digitaalisesta yhteiskunnasta ("Digital Agenda"; EU 2010), jonka kautta vaikutetaan verkkosisältöjen sisämarkkinoiden luontiin ja julkisen sektorin tietojen uudelleenkäytön toteutumiseen. Konkreettisena askeleena on myös EU:n tuottaman datan avoimuuden edistäminen, jotta sitä voidaan liittää suomalaisiin datakatalogeihin ja verkkopalveluihin.

^[8] Aineiston yhteistuotantomallia voisi pilotoida esimerkiksi tarjoamalla Liikenneviraston Digiroad-aineistoa OpenStreetMap-yhteisön jalostettavaksi.

Liite 1: Avoin julistus Euroopan julkisista palveluista

Nykypäivän yhteiskunnalliset haasteet ovat liian monimutkaisia yksinomaan hallinnon ratkaistaviksi. Siinä missä perinteiset hallintomallit ovat pyrkineet automatisoimaan julkisia palveluita ja ovat edistäneet itsepalvelua, tulee netin suurin vaikutus julkishallinnon kannalta olemaan yhteistyön, läpinäkyvyyden ja tiedon jakamisen vaikutus palveluiden kehittämiseen.

Euroopan tulisi tarttua mahdollisuuteen rakentaa kansalaisten ja valtion suhde uudelle perustalle, avaamalla instituutioita ja voimaannuttamalla kansalaiset olemaan aktiivisesti mukana julkisten palveluiden tuottamisessa.

Kansalaisina haluamme täyden näkyvyyden niihin toimiin, joita puolestamme tehdään. Haluamme, osallistua päätöksentekoon ja käytäntöjen kehittämiseen olemalla mukana päätösten valmistelussa, toimeenpanossa ja arvioinnissa. Haluamme olla aktiivisesti mukana julkisten palveluiden suunnittelussa ja tuotannossa. Tuomme esiin näkemyksemme ja haluamme säilyttää osan päätäntävallasta itsellämme. Haluamme koko julkisen informaation kirjon, lakiehdotuksista budjetti-informaatioon, helposti kansalaisten saataville, ymmärrettävässä, uudelleenkäytettävässä ja jalostettavassa muodossa. Tämä ei johdu siitä, että haluaisimme pienentää valtion roolia, vaan siitä että avoin yhteistyö parantaa julkisia palveluita ja päätöksenteon laatua.

Edelliseen tukeutuen ehdotamme kolmea perusperiaatetta Euroopan julkisille palveluille:

1. Läpinäkyvyys: kaikkien julkisen sektorin organisaatioiden tulisi olla “oletusarvoisesti läpinäkyviä”. Niiden tulisi tarjota yleisölle selkeää ja säännöllisesti päivittyvää informaatiota kaikista toimistaan ja päätöksentekoprosesseistaan. Kansalaisille tulisi taata vakituiset mekanismit läpinäkyvyyden puutteiden osoittamiseen. Informaatiota tarjotessaan julkisen sektorin organisaatioiden tulisi tehdä se avoimia, standardien mukaisia ja uudelleenkäytettäviä tiedostomuotoja hyödyntäen (luonnollisesti yksilöiden tietosuojan huomioiden).

2. Osallistuminen: julkishallinnon tulisi oma-aloitteisesti edistää kansalaisten osallistumista kaikkiin toimiinsa, kuten esimerkiksi palveluiden suunnitteluun ja lakien säätämiseen. Kansalaisten tuotoksien tulisi olla julkisesti muiden nähtävillä. Julkishallinnon tulisi vastata tuotoksiin julkisesti. Kansalaistyöteistyötaitojen tulisi olla hallinnon keskeinen osaamisalue.

3. Voimaannuttaminen: julkis-instituutioiden tulisi toimia mahdollistavina yhteisen hyvän luomisaluistoina. Hallinnollinen tieto ja hallinnon palvelut tulisi tarjota muiden käyttöön niin, että niiden varaan voidaan helposti rakentaa uutta. Julkishallinnon organisaatioiden tulisi tukea kansalaisten itsenäistä ongelmanratkaisua tarjoamalla välineitä, osaamista ja resursseja heidän käyttöönsä. Hallinnon organisaatioiden tulisi nähdä kansalaiset itseään koskevan tiedon omistajina ja taata tiedon käytön seuraaminen ja hallinta.

Ymmärrämme, että näiden periaatteiden toimeenpano vaatii aikaa ja resursseja, sillä hallinnon toimintatapoja joudutaan muuttamaan. Silti periaatteiden tulisi olla hallinnon kehittämisen keskiössä. Kansalaiset edistävät jo nykyään näitä periaatteita pyrkien julkisten palveluiden muuttamiseen “ulkopuolelta” käsin. Hallinnon tulisi tukea ja vauhdittaa tätä prosessia.

Vetoamme Euroopan valtioihin ja Euroopan komissioon, sen puolesta että nämä periaatteet otettaisiin mukaan sähköisen hallinnon toimeenpanosuunnitelmiin ja taattaisiin eurooppalaisille mahdollisuus nauttivat läpinäkyvästä, osallistuvasta ja voimaannuttavasta hallinnosta niin pian kuin mahdollista.

Liite 2: Kansalaisosallistujan työkalut osallistujat

Otso Kassinen	Kaavoituksen visualisointi: Palvelussa voi seurata visuaalisesti, millaisia rakennuksia kuntaan/kaupunkiin kaavoitetaan, ja vaikuttaa oman asuinympäristön muutoksiin.
Joseph Peters	SpeakUpForChange: The app brings the concept of deliberative democracy to social media on a major public policy issue.
Ilkka Törmä	YmpäristöYmpäristönsuunnittelutiedon lisääminen YTV:n reittioppaan kartalle levittäisi tietoa kaupunkien kehittymisestä laajalle yleisölle.
Kurt Linderoos	Personal EU: One person with same challenges from each EU country.
Samuel Rinnetmäki	Tietopaikka tarjoaa karttanäkymän tietoon, jolla on sijainti. Palvelu yhdistää useiden lähteiden tietoja ja vaikutuskanavia.
Marko Seppänen	Predikaatti+adjektiiviparit: Maailma kuvailtuna predikaatti+adjektiiviparien avulla, visualisoituna Google Mapsin päällä ja täydennettynä tietolähteillä.
Samuli Sairanen	Eduskunnan vahtikoira: Kansanedustajien sumeilematon tilastovusto
Tapio Nurminen	Nomen est omen: Verkkosovellus, jonka avulla voi analysoida suomalaisten sukunimien yleisyyttä, alkuperää ja sosiaalista asemaa.
Saara Paavola	Asialista: Kunnallishallintoon keskittyvä sovellus, jonka avulla kuntalaiset voivat osallistua oikea-aikaisesti päätöksentekoon.
Virpi Oinonen	The www.ydinvoima.fi -concept aims to facilitate public debate on a political issue by making MPs views on the issue easily accessible.
Katazina Andrukonyte	I am voting for change of enviromental awareness and present GoGreen tool, that tells about the pollution level in your living area.
Petri Louhelainen	Heti Kotiin - paikkatiedolla varustettu mobiilireittiopas kertoo nopeimman tien kotiin.
Harri Välimäki	Valtakunnallinen kyselyportaali, johon vastaajat on rekisteröitymisen yhteydessä todennettu pankkitunnusten avulla.
Tomi Terentjeff	YTV-reittiwidgetillä näet seuraavat joukkoliikenneyhteydet suoraan työpöydältäsi.
Kai Kousa	Uniform interface for accessing public data and creating applications of that data. Makes creating apps for these competitions a breeze.
Tomi Terentjeff	Hilmappi yhdistää HILMA-hankintailmoitukset OpenStreetMapsiin. Palvelu on suunnattu yrityksille helpottamaan HILMAN selausta.
Dani Pärnänen	Legiskooppi jakaa tietoa eduskunnassa säädettävistä laeista sekä edustajien ja puolueiden sopivuudesta käyttäjien kantoihin nähden.
Peter Tattersall	The Tax Tree is a proposal for a website that makes public sector expenditure visible to the public.
Antti Virolainen ja Juho Makkonen	Reittihjemuistutukset puhelimeesi automaattisesti juuri oikealla hetkellä, kun on aika lähteä kohti seuraavaa kalenteritapahtumaa.
Sami Majaniemi	Simupedia on palvelu, jossa yhteiskunnallisiin ongelmiin haetaan ratkaisuja yhteistyössä käyttäjien kesken simulaatioavusteisesti.
Antti Inkeroinen	Tavoitteena on luoda palvelu, jossa ihmiset pääsevät kertomaan mielipiteensä ja ottamaan kantaa heihin vaikuttaviin päätöksiin.
Janne Käki	Reittimarkus näyttää kartalla Helsingin seudun julkisen liikenteen ajoneuvojen liikkeitä siten, kuin ne aikataulun mukaan menisivät.
Ilari Kajaste	Palveluhakemisto kerää yhteen Helsingin alueen paikallisista palveluista sijainnin ja muita perustietoja



Kuva 1.1: Julkishallinnon datavarantoja ryhmiteltynä sisällön mukaisesti (Valtionhallinnon tietoarkkitehtuurit)

Liite 4: Lisensointimalleja

Creative Commons on helppokäyttöinen malli vastata tiedon käyttäjien epävarmuuteen siitä, mitä he saavat tehdä jollakin tietyllä verkosta löytyvällä aineistolla syyllistymättä tekijänoikeuksien loukkaukseen. Internetin mahdollistama helppo pääsy tiedon lähteille ja digitaalisen informaation helppo kopioitavuus vaativat myös helposti tunnistettavia tapoja identifoida se, mitä aineistolla saa tehdä. Creative Commons nojautuukin positiiviseen mikä on sallittu -asetelmaan tavanomaisista copyright-klausuuleista tutun mikä on kielletty -listan sijasta.

Lisenssien valinnassa lisenssien yhteensopivuus on oleellista. Tiedon yhdistelijä saattaa yhdistellä tietoa ja koodia monista eri lähteistä ja antamalla sen käyttöön, lisensoi sitä jollakin tavoin eteenpäin. Koodipuolella lisenssien yhdistely on tuottanut paljon päänvaivaa ja lisenssien hivuttamista yhteensopivaan muotoon. Näistä syistä täysin uusien lisenssien luomista Suomeen tulee välttää. Esimerkiksi Wikipediassa on hiljattain siirrytty Creative Commons Attribution-ShareAlike -lisenssiin. Lisensoinnin valintaa täytyy myös miettiä, kun tietoa otetaan vastaan verkon käyttäjiltä (molemmipuolinen Creative Commons voi olla reilu valinta). Whitehouse.gov on Barak Obaman hallintokauden alussa velvoittanut kolmannet osapuolet myöntämään sivustolla julkaistun sisällön Creative Commons -lisenssillä (<http://creativecommons.org/weblog/entry/13591>). Näin sivuston käyttäjä voi tuttuihin lisenssikäytäntöihin nojautuen lainata toisen kansalaisen sivustolla julkaisemaa materiaalia, esimerkiksi hallinnolle kohdis-

tettua ongelman kuvausta, osana omaa Creative Commons sisältöään.

Creative Commons -lisenssin ongelma on Noncommercial-attribuutin monitulkintaisuus. On olemassa selkästi kaupallista ja selkeästi ei-kaupallista toimintaa. Mutta on myös olemassa epäselvää monitulkintaista toimintaa suhteessa sisältöjen ja tiedon julkaisemiseen. Kierrätettyä Creative Commons -sisältöä julkaiseva taho saattaa esimerkiksi saada voittoa, kattaa kulujaan tai saada kilpailuetua epäsuorilla mainos- tai koulutustuloilla. Epäkaupallista Creative Commonsia on käsitelty oikeuslaitoksissa melko vähän. Sen juridiseen tulkintaan liittyy siksi epävarmuutta. Noncommercial-lisenssin käyttäjän kannattaa arvioida omia vapaan lisensoimisen tavoitteitaan ja punnita avaamia mahdollisuuksia suhteessa haittoihin ja riskeihin. Arvioimista vaikeuttaa se, että on vaikea ennakoida kuinka hyvin lisensoitu sisältö ylipäättään tulee leviämään, sillä tilanteeseen vaikuttaa käyttäjästä itsestään riippumattomia ja vaikeasti ennakoitavia tekijöitä.

Liite 5: Palvelukeskeinen arkkitehtuuri

Palvelukeskeinen arkkitehtuuri (Service Oriented Architecture, SOA) on suunnittelutapa, jonka avulla eri tietojärjestelmien toiminnot ja prosessit on suunniteltu toimimaan itsenäisinä, avoimina ja joustavina palveluina. Myös julkishallinnossa palvelukeskeisten arkkitehtuurien nähdään olevan keskeinen lähestymistapa uusien verkkopalveluiden kehityksessä. Monessa SOA onkin kertymä hyviä suunnittelukäytäntöjä, kuten että palveluiden hyödyntäminen perustuu aina avoimien (standardien tai muutoin vakioitujen) rajapintojen ja tekniikoiden käyttöön. Palvelukeskeistä arkkitehtuuria voidaan toteuttaa useilla eri teknologioilla ja teknologisista valinnoista palveluiden rajapintojen toteutustapa on keskeisin. Rajapintojen toteutustapoina tässä esitellään järeämpi yritysten välisiin palveluihin tyypillisesti käytetty web-palvelutekniikka (Web Services) ja ketterämpi avoimen datan julkaisuun hyvin soveltuva REST-tyyli (Representational State Transfer).

Web-palvelutekniikka (Web Services)

Web-palvelutekniikka on toistaiseksi kaikkein yleisimmin käytetty palvelukeskeisten arkkitehtuurien toteutustapa ja SOA:n yleistymisen liittykin paljolti juuri web-palvelutekniikan yleistymiseen. Viestien välitykseen järjestelmien välillä käytetään web-palveluissa kahta protokollaa HTTP ja SOAP ja web-palvelutekniikan mukaisia rajapintoja kutsutaankin usein SOAP-palveluiksi. Web-palvelutekniikalla toteutetun rajapinnan kautta välitetään tietosisältö aina XML-muodossa standardin «kehysrakenteen» sisällä. Näin ollen web-palvelutekniikan rajapintojen kautta ei saa suoraan välitettyä esimerkiksi RSS-syötettä, vaan kehysrakenne pitää purkaa (kirjekuori pitää avata) vastaanottajan päässä. Web-palveluiden (Web Service) toteutus ei vielä tarkoita, että tehtäisiin palvelukeskeisen arkkitehtuurin mukaisia palveluita. Myöskään SOA:n toteutus ei vaadi web-palvelutekniikan käyttöä, vaan toteutusteknologiana voi olla jokin muukin, kuten seuraavaksi esiteltävä REST (Representational State Transfer).

REST

Viimevuosina nettiohjelmoijien suosioon noussut REST (Representational State Transfer) on arkkitehtoninen tyyli hajautettuja järjestelmiä varten. Yksi alkuperäisistä HTTP-protokollan pääkirjoittajista Roy Fielding esitteli tämän arkkitehtuurityylin väitöskirjassa vuonna 2000. REST on yksi tapa toteuttaa samankaltaista toiminnallisuutta, kuin mihin päästään web-palvelutekniikalla, mutta käyttämällä vain HTTP-protokollaa. REST-ajattelu sopii erityisen hyvin datan julkaisemiseen ja koosteisten palvelujen (mashupit) taustalle, sillä siinä keskeinen tekijä on itse sisältö (englannissa usein resource). Aivan, kuten web-arkkitehtuurissakin, niin REST-tyylissä kaikki palvelut nähdään resursseina, joilla on yksi-

loitävissä oleva URI-tunniste ja yhtenäinen standardi HTTP-protokollan mukainen kutsurajapinta. Www itsessään on esimerkki REST-periaatteella toimivasta palvelusta. Suuret verkkopalveluiden tuottajat, kuten Amazon, Yahoo, Google ja Facebook suosittelevat jo julkaisemiensa helppokäyttöisten REST-rajapintojen käyttöä. REST-palvelujen dokumentaatio on usein kevyttä: Tässä ovat osoitteet ohjelmointirajapintamme luonti-, haku-, muokkaus- ja poisto-operaatioihin. Niille voi välittää tällaisia parametreja. Ole hyvä ja käytä.

REST ei ole edellä kuvattu web-palvelu (Web Service HTTP + SOAP), eikä Remote Procedure Call (RPC) malli eikä Web-arkkitehtuuri.

Vaatii poisoppimista vanhoista käytännöistä, tavoista ja kulttuurista, esimerkiksi RPC on pesiytynyt yllättävän syvälle kaikkeen mikä liittyy hajautettuihin järjestelmiin

Lähdeluettelo

Benson, Yrjö (2009). Valtion IT-toiminta. Valtion IT-johtaja Yrjö Benson alustus Kuntaliiton seminaarissa 4.11.2009. Helsinki: Kuntaliitto. qsb.webcast.fi/k/kuntaliitto/kuntaliitto_2009_1104_07_benson/. Viitattu 10.3.2010.

Berners-Lee, Tim (2006). Linked Data. www.w3.org/DesignIssues/LinkedData.html. Viitattu 10.3.2010

Cloud, John (2006). The Gurus of YouTube. 16.12.2006. Time magazine. www.time.com/time/printout/0,8816,1570721,00.html. Viitattu 8.3.2010.

DCMS, & BIS (2009). Digital Britain. Final report. Building Britain's future. DCMS (Department of culture, media and sport) and BIS (Department of Business Innovations and Skills). www.culture.gov.uk/images/publications/digitalbritain-finalreport-jun09.pdf. Viitattu 15.12.2009.

Dekkers, Makx, Polman, Femke, de Velde, Robbin, & de Vries, Marc (2006). MEPSIR Measuring European Public Sector Information Resources. Final Report of Study on Exploitation of public sector information – benchmarking of EU framework conditions. ec.europa.eu/information_society/policy/psi/docs/pdfs/mepsir/final_report.pdf. Viitattu 10.3.2010.

EU (2007). Lausunto 4/2007 henkilötietojen käsitteestä. 20.6.2007. ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2007/wp136_fi.pdf. Viitattu 8.3.2010.

EU (2007b). Communication from the Commission. Pre-commercial Procurement: Driving innovation to ensure sustainable high quality public services in Europe. SEC (2008) 1668. http://ec.europa.eu/invest-in-research/pdf/download_en/com_2007_799.pdf Viitattu 17.3.2010

EU (2009). Cellphone charger harmonization. ec.europa.eu/enterprise/sectors/rtte/chargers/index_en.htm. Viitattu 8.3.2010.

EU (2010). EUROPE 2020 A strategy for smart, sustainable and inclusive growth. <http://ec.europa.eu/eu2020/pdf/COMPLET%20EN%20BARROSO%20%20%20007%20-%20Europe%202020%20-%20EN%20version.pdf> Viitattu 15.3.2010

Eups20 (2010). An Open Declaration on European Public Services. <http://eups20.wordpress.com/the-open-declaration/> Viitattu 18.3.2010

EY Komissio (2009). Julkisen sektorin hallussa olevien tietojen uudelleenkäyttö - direktiivin 2003/98/EY uudelleen tarkastelu. Komission tiedonanto KOM(2009) 212 lopullinen. www.eipsiplatform.eu/media/files/com09_212_fi. Viitattu 5.2.2010.

EY (2003). Euroopan parlamentin ja neuvoston direktiivi 2003/98/EY, annettu 17. päivänä

marraskuuta 2003, julkisen sektorin hallussa olevien tietojen uudelleenkäytöstä. www.diges.info/pdf/PSI-direktiivi-suomi.pdf. Viitattu 8.3.2010.

Forss, Marko (2009). Virtuaalinen lähipoliisiryhmä. Poliisin kotisivut. www.poliisi.fi/irc-galleria. Viitattu 8.3.2010.

García-García, J., & Alonso de Magdaleno, I. (2010). Commons-based innovation, The Linux kernel case. In 2nd European Conference on Corporate R&D. iri.jrc.ec.europa.eu/concord-2010/posters/Garcia-Garcia.ppt. Viitattu 9.3.2010.

HS (2010). Ministeriö pohtii viranomaistietojen vapauttamista verkkoon. Kansalaisista virkatiedon jalostajia. Lehtiartikkeli 7.2.2010. Helsingin Sanomat.

HS (2010). Suomi elää velaksi 2010. Lehtiartikkeli 9.1.2010. Helsingin Sanomat.

Hermans, Outi, & Hermans, Raine (2009). Paikkatietojen yhteiskäyttö ja jakeluperiaatteet. Hinnoitteluperiaatteiden analyysi ja kansantaloudellisten vaikutusten simulointi. Keskusteluaiheita – 1194. www.etla.fi/files/2372_Dp1194.pdf. Viitattu 20.12.2009.

Hintikka, Kari A. (1993). Tieto - neljäs tuotannontekijä : tehtaasta televirtuaalisuuteen. Helsinki: Painatuskeskus.

Hintikka, Kari A. (2009). Toimintatapojen muutoksia eli kollektiiviset ja hajautetut toimintamallit. Alustus Suomi.fi:n seminaarissa 7.4.2009. www.slideshare.net/ubiq/suomifi-seminaari-742009-alustus-kari-a-hintikka. Viitattu 7.3.2010.

Howe, Jeff (2006). The Rise of Crowdsourcing. Wired magazine. www.wired.com/wired/archive/14.06/crowds.html. Viitattu 8.3.2010.

Joey Van Den Hurk, Peter Rem, M. J. (2008). Standard Cost Model for Citizens - User's guide for measuring administrative burdens for Citizens. Ministry of the Interior and kingdom Relations. whatarelief.eu/publications/standard-cost-model. Viitattu 10.3.2010

Kansallinen ennakointiverkosto (2008). Internet ja vuorovaikutuksen uudet muodot. (Nieminen-Sundell, Riitta, toim.). [www.foresight.fi/wp-content/uploads/2009/08/Internet ja vuorovaikutuksen uudet muodot.pdf](http://www.foresight.fi/wp-content/uploads/2009/08/Internet_ja_vuorovaikutuksen_uudet_muodot.pdf). Viitattu 19.1.2010.

Karvonen, Erkki (2000). Elämmekö tieto- vai informaatioyhteiskunnassa? Teoksessa Vuorensyrjä, Matti & Savolainen, Reijo (toim.) Tieto ja tietoyhteiskunta. Helsinki: Gaudeamus

Kuntaliitto (2010). Kuntien verkkoviestintäohje. Kuntaliiton verkkojulkaisu. Kuntaliitto. hosted.kuntaliitto.fi/intra/julkaisut/pdf/p20100127150612668.pdf. Viitattu 10.3.2010.

Kuronen, Timo (1998). Tietovarantojen hyödyntäminen ja demokratia. Helsinki: Suomen itsenäisyyden juhlarahasto. www.sitra.fi/julkaisut/tietoyhteiskunta/sitra174.pdf?download=Lataa+pdf. Viitattu 12.12.2009.

Kuronen, Timo (1998b). Tietovarantojen hyödyntäminen ja demokratia. Esimerkkejä tiedon prosesseista. SITRA 174. Helsinki: Suomen itsenäisyyden juhlarahasto. onko tähän viitattu? -Antti "Jogi" Poikola 10/03/2010 19:51

LVM, & VM (2010). Työryhmät edistämään julkisen tiedon saatavuutta. Tiedote 11.02.2010. Valtiovarainministerio ja Liikenne- ja viestintäministeriö. www.lvm.fi/web/fi/tiedote/view/1130523. Viitattu 8.3.2010.

LVM (2009). Sähköisesti nouseva Suomi Viestinnän elinkeinopoliittisen työryhmän loppuraportti 22.10.2009. www.lvm.fi/c/document_library/get_file?folderId=339549&name=DLFE-9534.pdf. Viitattu 3.3.2010.

Lasica, Joseph Daniel (2009). Identity in the Age of Cloud Computing. www.aspeninstitute.org/sites/default/files/content/docs/pubs/Identity_in_the_Age_of_Cloud_Computing.pdf. Viitattu 5.3.2010.

Luoto, Karoliina (2009). Suomen tietoyhteiskunta 2020. Foresight.fi -blogimerkintä

17.12.2009. Sitra. www.foresight.fi/2009/12/17/suomen-tietoyhteiskunta-2020/. Viitattu 15.2.2010.

Mannermaa, Mika (2006). Demokratia tulevaisuuden myllerryksessä. Yhteiskunnallinen vaikuttaminen uudessa viitekehyksessä. Helsinki, Eduskunta, tulevaisuusvaliokunta. www.eduskunta.fi. Viitattu 10.3.2010. viitataan tähän? -Antti "Jogi" Poikola 10/03/2010 20:07

MMM (2004). Kansallinen paikkatietostrategia 2005–2010, Maa- ja metsätalousministeriö MMM:n julkaisuja 10/2004. Retrieved from www.mmm.fi/fi/index/etusivu/maanmittaus_paikkatiedot/paikkatietojenyhteiskaytto/kansallinenpaikkatietostrategia.html. Viitattu 10.3.2010.

Mokka, Roope & Neuvonen, Aleksi (2006). Yksilön ääni. Hyvinvointivaltio yhteisöjen ajalla. Helsinki: Sitra. www.sitra.fi/julkaisut/raportti69.pdf. Viitattu 8.3.2010.

Neuvottelukunta (2009). Tietoyhteiskunnan vauhdittamiseen löydettävä ratkaisuja, Lehdistötiedote 24.11.2009. www.lvm.fi/web/fi/tiedote/view/994565. Viitattu 10.3.2010.

Niiniluoto, Ilkka (1989). Informaatio, tieto ja yhteiskunta. Filosofinen käsiteanalyysi. Helsinki: Valtionhallinnon kehittämiskeskus.

Niiniluoto, Ilkka (2000). Uskomuksia ilman tietoa? www.uta.fi/~attove/niini.pdf. Viitattu 7.3.2010.

O'Reilly, Tim (2010). Open Government: Collaboration, Transparency, and Participation in Practice. (D. Lathrop & L. Ruma) (p. 432). O'Reilly Media; 1 edition.

OECD (2006). OECD Principles and Guidelines for Access to Research Data from Public Funding. www.oecd.org/dataoecd/9/61/38500813.pdf. Viitattu 20.1.2010.

OECD (2008). OECD Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information. www.oecd.org/dataoecd/0/27/40826024.pdf. Viitattu 10.3.2010.

Open Knowledge Foundation (2006). Open Knowledge Definition (OKD). Open Knowledge Foundation. www.opendefinition.org/okd/. Viitattu 10.3.2010.

Otakantaa.fi (2010). Miten suomalaista tietoyhteiskuntaa pitäisi kehittää? Liikenne- ja viestintäministeriön kysely Otakantaa.fi -palvelussa. www.otakantaa.fi/forum.cfm?group=428&sort=Otsikko&id=&ref=Kaikki keskustelut&pg=4. Viitattu 10.3.2010

Pauku, Paul (2009). Julkiset sähköiset palvelut / kehittämisen pullonkaulat. Selvitysmiehen raportti 12.6.2009. www.lvm.fi/web/fi/julkaisu/view/906227. Viitattu 7.3.2010.

Pursiainen, Harri (2009). Kansallinen älyliikenteen strategia. Selvitysmiehen ehdotus 30.10.2009. www.lvm.fi/web/fi/tyoryhmat/tyoryhma/view/845060. Viitattu 10.3.2010.

Shirky, Clay (2008). Here Comes Everybody: The Power of Organizing Without Organizations. Penguin Press HC.

TEM (2008). Kansallinen innovaatiostrategia. http://www.tem.fi/files/19704/Kansallinen_innovaatiostrategia_12062008.pdf Viitattu 15.3.2010

TEM (2009). Kilpailulaki 2010 -työryhmän mietintö. www.tem.fi/files/21617/TEM4.pdf. Viitattu 10.3.2010

TEM (2010). Kysyntä- ja käyttäjälähtöisen innovaatiopolitiikan toimenpideohjelma. www.tem.fi/files/26093/OSA_2_final.pdf. Viitattu 20.2.2010.

Torkington, Nat (2010). Rethinking Open Data. Lessons learned from the Open Data front lines. O'Reilly Radar blog. radar.oreilly.com/2010/02/rethinking-open-data.html. Viitattu 20.1.2010.

Turkki, Teppo (2009). Nykyaikaa etsimässä. Suomen digitaalinen tulevaisuus. www.eva.fi/files/2573_nykyaikaa_etsimassa.pdf. Viitattu 8.3.2010.

UK (2009). Putting the Frontline First: smarter government. www.hmg.gov.uk/media/52788/smarter-government-final.pdf. Viitattu 8.3.2010

VM (2003). Hallinnon sisäisten tietoluovutusten hinnoittelu, Hallinnon tietoluovutusten hinnoitteluhankkeen raportti 30.5.2003. Työryhmämuistioita 16/2003. www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/01_julkaisut/04_hallinnon_kehittaminen/37893/40524_fi.pdf. Viitattu 26.1.2010.

VM (2004). Hallinnon sisäisten tieto- luovutusten maksukäytäntöjä selvittäneen työryhmän muistio. Työryhmämuistioita 11/2004. www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/01_julkaisut/04_hallinnon_kehittaminen/89111_fi.pdf. Viitattu 5.3.2010.

VM (2005). Direktiivi 2003/98/EY julkisen sektorin hallussa olevien tietojen uudelleenkäytöstä. Kansallinen implementointi ja vastaavuus Suomen lainsäädäntöön. Muistio 26.5.2005. www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/03_muut_asiakirjat/95083.pdf. Viitattu 8.3.2010.

VM (2009). Valtiotason arkkitehtuurit -hanke. Julkishallinnon tietoarkkitehtuuri Määrittely 0.3 21.12.2009. www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/03_muut_asiakirjat/20100105Valtio/04_Julkishallinnon_tietoarkkitehtuuri_v0.3.pdf. Viitattu 15.1.2010.

VM (2009b). Sähköisen asioinnin ja demokratian vauhdittamisohjelman (SADe) toteuttamissuunnitelma 2009-2014. Valtiovarainministeriön muistio 16.6.2009. www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/03_muut_asiakirjat/suunnitelma_SADe_160609.pdf. Viitattu 10.3.2010

VM (2009c). Valtiotason arkkitehtuurit -hanke. Valtionhallinnon arkkitehtuuriperiaatteet ja -linjaukset. Määrittely 0.62 Päiväys 31.12.2009. http://www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/03_muut_asiakirjat/20100105Valtio/01_Valtionhallinnon_arkkitehtuuriperiaatteet_ja_linjaukset_v0.62.pdf Viitattu 15.3.2010

VM (2010). Tilastotoimen tehostamista ja alueellistamista valmistelleen työryhmän loppuraportti. http://www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/01_julkaisut/04_hallinnon_kehittaminen/20100302Tilast/Loppuraportti_2_3_2010_liitteet.pdf Viitattu 10.3.2010.

VNK (2006). Kansallinen tietoyhteiskuntastrategia 2007-2015. Uudistuva, ihmisläheinen ja kilpailukykyinen Suomi. www.arjentietoyhteiskunta.fi/files/34/Kansallinen_tietoyhteiskuntastrategia.pdf. Viitattu 10.3.2010.

Voutilainen, Tomi (2009). ICT-oikeus sähköisessä hallinnossa. Väitöskirja Joensuun Yliopisto (p. 360). Edita Publishing Oy.

W3C (2004). Architecture of the World Wide Web, Volume One W3C Recommendation 15 December 2004. W3C. www.w3.org/TR/webarch/. Viitattu 10.3.2010

YK (2010). Green Growth. UNESCAP. www.greengrowth.org/.

YLE (2008). Yleisradion hallituksen toimintakertomus v. 2008. yle.fi/yleista/kuvat/2008tilinpaatos.pdf.

YLE (2009). Tietosuojavaltuutettu vaatii yleistä tietoturvalakia. 20.2.2009. YLE Uutiset verkkosivu. yle.fi/uutiset/kotimaa/2009/02/tietosuojavaltuutettu_vaatii_yleista_tietoturvalakia_560232.html. Viitattu 8.3.2010.

YLE (2009). Ylen mahdollistajastrategia. Yle. yle.fi/yleista/kuvat/YLE_mahdollistajastrategia.pdf. Viitattu 8.3.2010. yle.fi/yleista/kuvat/YLE_mahdollistajastrategia.pdf. Viitattu 8.3.2010.

YLE (2009). Ylen mahdollistajastrategia. Yle. http://yle.fi/yleista/kuvat/YLE_mahdollistajastrategia.pdf. Viitattu 8.3.2010.

