



24.6.2014

kirjaamo@lvm.fi
emil.asp@lvm.fi

Lausuntopyyntö LVM/138/05/2014

VÄESTÖREKISTERIKESKUKSEN LAUSUNTO KOSKIEN KANSALLISTA BIG DATA-STRATEGIAA

Väestörekisterikeskus lausuu pyynnöstä seuraavaa:

Yleistä

Väestörekisterikeskuksen näkemyksen mukaan kansallisen Big Data –strategian valmistelu on kannatettavaa ja Väestörekisterikeskus pitää hyvänä, että selvitetään yleisiä edellytyksiä big datan hyödyntämiselle ja sitä kautta tietoaaineistojen ja uusien analyysimenetelmien laajempaa hyödyntämistä yhteiskunnan eri sektoreilla.

Väestörekisterikeskus katsoo, että Väestörekisterikeskuksen ja muiden perusrekisterien ylläpitäjien tulisi olla aktiivisesti mukana big dataa koskevassa työssä. Suomessa on erittäin laadukkaita perusrekistereitä; väestötietojärjestelmä, yritys- ja yhteisötietojärjestelmä ja kiinteistö- ja rakennustietojärjestelmä, jotka muodostavat yhteiskunnallisesti merkittävän tietokokonaisuuden, jota voidaan hyödyntää ja joita jo nyt hyödynnetään erilaisten big data -menetelmien aineistona.

Strategialuonnoksessa mainittuihin big datan edelläkävijävaltioihin, Yhdysvaltoihin ja Iso-Britanniaan, verrattuna Suomessa käytössä olevat perusrekisterit ovat poikkeuksellisen korkealuokkaisia laadultaan ja kattavuudeltaan ja niiden tarjoamalle perustalle on mahdollista rakentaa myös erittäin kehittyntä big data -toimintaa.

Strategialuonnoksessa esitetyt toimenpide-ehdotukset

Väestörekisterikeskus haluaa kiinnittää huomiota seuraaviin strategialuonnoksessa esitettyjen toimenpideluonnosten kohtiin ja niihin liittyviin näkökulmiin:

6.1. Datatietoisuus

Erityisesti liittyen toimenpide-ehdotukseen ”b” nyt olemassa olevaa perusrekisterien asiantuntijuutta tulisi hyödyntää mahdollisimman paljon big data-klusterin kehittämisessä.





24.6.2014

Jos big data –menetelmiä ryhdytään rakentamaan perusrekisterien päälle, olisi varmistettava, että toiminnassa mukana olevilla tahoilla on riittävä ymmärrys rekisterien tietosisällöstä ja ylläpitomenettelyistä. Näin vältetään tilanne, että rekisterien kaikkia hyötyjä ei saataisi realisoitua, koska käyttäjät eivät tunne niiden sisältöä. Tärkeää on myös, että datan käyttäjät tiedostavat lainsäädännön asettamat rajoitukset datan käytölle, koska mainittuihin perusrekistereihin kohdistuu niitä koskevaa erityislainsäädäntöä.

6.7 Datan avoimuus ja käytettävyys

Kaikki toimenpide-ehdotuksen alakohdat (p,q,r,s) ovat sinänsä kannatettavia. Väestörekisterikeskus haluaa kuitenkin kiinnittää huomiota datan avoimuuteen liittyvän anonymiteetin varmistamiseen.

On tärkeää huomioida, kuten strategialuonnoksessa on tehty, että kaikki big data ei ole avointa dataa (mm. kappale 2.7, s.16), vaan suuri osa nyt syntyvästä big data -aineistosta koskee henkilöitä, jolloin sen käyttöä säätelee henkilötietolaki. Henkilötietolain alaisia henkilötietoja ei puolestaan voida käsitellä avoimena tietona, vaan ne tulisi anonymisoida ennen kuin aineistoa voitaisiin käsitellä avoimena tietona. Tämä anonymisointi ei kuitenkaan ole täysin ongelmaton, koska big datan ja avoimen tiedon luonteeseen kuuluu sen yhdistäminen eri lähteistä saatavaan tietoon. Tästä seuraa se, että yhdessä tietolähteessä toteutettu tiedon anonymisointi ei vielä takaa sen pysymistä anonymina. Väestörekisterikeskus suhtautuu tästä syystä myös anonymisoidun big data –aineiston käyttöön hyvin varovaisesti.

6.8 Sääntely

Toimenpide-ehdotukseen ”t” liittyen Väestörekisterikeskus haluaa korostaa, että tehtäessä selvitystä big dataan vaikuttavista laeista, tulee huomioida myös laki väestötietojärjestelmästä ja Väestörekisterikeskuksen varmennepalveluista (661/2009), koska jos big data –menetelmiä rakennetaan väestötietojärjestelmästä saatavan datan varaan, niiden käyttöä säätelee myös yllä mainittu laki.

Liittyen toimenpide-ehdotukseen ”v” osana big dataa -koskevan sääntelyn tai etikeitin kehittämistä tulisi kiinnittää erityistä huomiota siihen, mikä on hyväksyttävä henkilöiden seuraamisen ja profiloinnin taso ja mikä olisi siltä edellytettävä läpinäkyvyyden aste. Strategialuonnos nostaa uhkana esille mm. liiallisen profiloinnin seurauksena tapahtuvan diskriminoinnin (esim. kappale 1.2, s.4). Tämän uhan torjumiseksi olisi säädettävä, mitä tietoja ei saisi käyttää profilointiperusteena (esim. etnistä alkuperää kuvaava tieto), koska muuten profilointiin käytetään kaikkea tietoa, jolla havaitaan olevan tilastollinen korrelaatio tutkittavan muuttujan kanssa, vaikka tämä saattaa olla diskriminoivia seurauksia. Jos tätä säätelystä ei tehdä keskitetysti, se jäänee yksittäisten viranomaisien tulkittavaksi, minkä seurauksena voi muodostua useita erilaisia tulkintoja siitä, mikä on sallittua ja mikä ei. Toinen tehokas tapa torjua diskriminoivia profilointiperusteita on se, että käytettyjen muuttujien tulee olla läpinäkyviä profiloinnin kohteelle, koska mahdollisten diskriminoivien perusteiden julki tuleminen ehkäisee niiden käyttöä.

Toinen pohdittava asia toimenpide-ehdotukseen ”v” ja profilointiin liittyen on se, kuinka vanhoja tietoja profiloinnissa on sallittua hyödyntää ja kuinka pitkään henkilöitä koskevia profiileja saa säilyttää. Monille rekisterimerkinnöille on lailla säädetty vanhenemisaika (esim. rikosrekisteri), mutta vaikka tämän tyyppinen rekisterimerkintä vanhenee, se saattaa silti jäädä elämään henkilöön kiinnitetyn profiilin kautta.



24.6.2014

On varmistettava, että big data –aineiston keruusta seuraavalla yksityiskohtaisemalla profiloinnilla on kansalaisten hyväksyntä ja että he eivät koe sen loukkaavan yksityisyydensuojaansa. Tämän varmistaminen on kuitenkin vaikeaa. Esimerkiksi lyhyellä tähtäimellä yksittäisen yrityksen on helppo saada hyväksyntä tietojen käyttöön, vaikkapa lupaamalla alennuksia vakuutusmaksuissa vastineeksi auton GPS-datasta, mutta eri tietolähteistä vähitellen syntyvän kokonaisuuden hahmottaminen on hankalaa, ja sen lopputulemaa voi olla vaikea peruuttaa, vaikkei sille enää haluaisikaan antaa hyväksyntäänsä.

Profilointien osalta on toki huomioitava myös valmisteilla olevan sääntely (tietosuoja-asetus) ja sen mahdolliset vaikutukset big data –ajatteluun ja tietojen mahdollisiin käyttötarkoituksiin.

6.11. MyData

Strategialuonnos toteaa, että ns. MyData -näkökulmilla voidaan purkaa big dataan ja yksityisyydensuojaan liittyviä uhkia (kappale 5.1, s.34). Antamalla henkilölle päätösvalta siitä, kuka hänen tietojansa saa käyttää ja mihin tarkoitukseen tuo kontrollin tietojen käytöstä kansalaisille itselleen ja sitä kautta torjuu aineiston väärinkäytön riskejä ja sen käyttöä henkilön tietämättä.

Yleisenä kommenttina MyData-ajatteluun liittyen Väestörekisterikeskus toteaa, että MyData pohdinnoissa on tärkeää huomioida se tosiasia, että julkisen vallan käytön yhteydessä henkilötietojen käsittely ei pääsääntöisesti perustu henkilön omaan suostumukseen, vaan siitä säädellään erikseen laissa (viranomaistoimintaa ja viranomaisten henkilörekistereitä koskevat toimialakohtaiset erityislait). Henkilön itsemääräämisoikeus viranomaisrekisterissä oleviin tietoihin nähden pitää tyypillisesti sisällään omien tietojen tarkastusoikeuden, oikeuden kieltää tietojen luovuttaminen tiettyihin tarkoituksiin sekä oikeuteen saada virheelliset tai muuten epätäydelliset tiedot korjattua. Pääsääntöisesti henkilö ei voi itse määrätä, rekisteröidäänkö hänen tietojansa viranomaisrekisteriin tai miten niitä voidaan luovuttaa.

MyData -ajattelun vahvistuminen tulee lisäämään keskustelua myös viranomaisten henkilörekistereissä olevien tietojen käsittelyn periaatteista ja muun muassa omien tietojen tarkastusmahdollisuuksien tekninen ja toiminnallinen kehittäminen on välttämätöntä ja kannatettavaa. Toimenpide-ehdotuksiin tulisi siitä syystä lisätä MyData mallien soveltuvuuden viranomaistoimintaan selvittäminen. Väestörekisterikeskus on halukas osallistumaan tätä koskevaan keskusteluun.

Big Dataan liittyvät toimenpiteet Väestörekisterikeskuksessa

Väestötietojärjestelmän tietoja käytetään jo tällä hetkellä tiettyssä mielessä big data –tyyppisesti. Esimerkiksi useat yritykset profiloivat asiakasrekistereitään väestötietojärjestelmän aineistoa vasten. Tämän profiloinnin toteuttavat Väestörekisterikeskuksen yhteistyökumppanit (Bisnode Marketing Oy, Fonecta Enterprise Solutions Oy, Itella Posti Oy), joilla on pääsy sekä yrityksen asiakasrekisteriin, että väestötietojärjestelmän tietoihin. Varsinaisilla profiilien hyödyntäjillä ei siis ole pääsyä niiden laitimiseen käytettävään aineistoon. Tässä toiminnassa Väestörekisterikeskus säätelee tarkasti sitä, mitä väestötietojärjestelmän tietoja profilointiin on sallittua käyttää ja mihin rekistereihin väestötietojärjestelmän tietoja on sallittua yhdistää. Datan käy-



24.6.2014

tölle on asetettu reunaehdot, joilla pyritään varmistamaan henkilöiden yksityisyydensuojan toteutuminen. Anonymisoituja väestötietojärjestelmän tietoja on käytetty myös erilaisten luotonmyöntöprofiilien rakentamisessa.

Väestötietojärjestelmän tietoja käytetään myös tutkimustyössä esimerkiksi sairauksien periytyvyyden tai sairauksien ja asumishistorian yhteyksien selvittämisessä. Tämä tietojen käyttö perustuu kuitenkin lakiin väestötietojärjestelmästä ja Väestörekisterikeskuksen varmennepalveluista, jonka nojalla väestötietojärjestelmän tietojen käyttö voi perustua vain Väestörekisterikeskuksen antamaan erilliseen tietolupaun, jossa määritellään mitä väestötietojärjestelmän tietoja organisaatio saa käyttöönsä ja mikä on tietojen sallittu käyttötarkoitus. Big data -menetelmien muun tyyppinen soveltaminen edellyttäisi muutoksia mainittuun lakiin.

Lakiin väestötietojärjestelmästä ja Väestörekisterikeskuksen varmennepalveluista tehdyn rakennustietojen koskevan muutoksen nojalla Väestörekisterikeskus tulee vuoden 2015 alusta kuitenkin julkaisemaan myöhemmin tarkemmin määriteltäviä rakennustietoaaineistoja avoimena tietona, sillä edellytyksellä, että valtionvarainministeriö kompensoi tietojen avaamisesta aiheutuvat tulonmenetyksen virastolle. Avattava rakennustietoaaineisto koskee koko maan rakennuskantaa ja se muodostaa merkittävän big data -aineiston, jota voidaan hyödyntää useilla eri tavoilla. Ennen datan avaamista on kuitenkin varmistettava, että avattava kokonaisuus on sellainen, että se ei täytä henkilötietolain määritelmää henkilötiedolle, eli ts. että se ei ole yhdistettävissä henkilöön ja kuvaa hänen elinolosuhteitaan henkilöä koskevalla tavalla.

Omanlaisensa big data -aineiston muodostavat myös väestötietojärjestelmän lokitiedot. Lokitiedoilla saattaisi olla tietyn tyyppistä big data -arvoa, jolla voitaisiin arvioida esimerkiksi sitä, mihin eri henkilön elämänvaiheisiin liittyy eniten viranomaisasiointia. Lainsäädäntö asettaa kuitenkin hyvin tiukat rajoitukset sille, mihin lokiaineistojen käyttö on sallittua, eikä se nykytilanteessa ole sovellettavissa big data -aineistoksi.

Väestörekisterikeskuksella ei ole varsinaisia tunnistettavia big data -aineistoja väestötietojärjestelmän tietojen (ja sen lokitietojen) lisäksi ja viraston varsinainen tehtävä on näiden tietojen ylläpito ja niiden luovuttamisen hallinnointi. Lisäksi nämä tiedot eivät sellaisenaan ole otettavissa big data -menettelyjen piiriin, vaan niiden käyttöä rajaa ja ohjaa laki väestötietojärjestelmästä ja Väestörekisterikeskuksen varmennepalveluista.

Väestörekisterikeskuksen rooli kansallisessa Big Data -kehittämistyössä

Kansallinen big data -kehittämistyö koskettaa monin tavoin Väestörekisterikeskuksen toimialaa, koska Väestörekisterikeskus on merkittävän kansallisen tietovarannon, väestötietojärjestelmän, rekisterinpitäjä. Big data -käytössä keskeinen lainsäädäntö (henkilötietolaki, laki väestötietojärjestelmästä ja Väestörekisterikeskuksen varmennepalveluista) ohjaa myös Väestörekisterikeskuksen toimintaa ja Väestörekisterikeskuksen tulisi olla mukana lainsäädännön mahdollisessa muutostyössä.



24.6.2014

Big data -aineistojen kestävä soveltaminen edellyttää tasapainon löytämistä yksityisyydensuojan ja yritysten tai yhteiskunnan tehokkuusvaatimusten välillä. Väestörekisterikeskuksella on runsaasti kokemusta näiden keskenään usein jännitteistenkin ulottuvuuksien yhteensovittamisesta. Jos ryhdytään kehittämään suomalaista big data -etikettiä, tulisi Väestörekisterikeskuksen olla mukana tässä valmistelutyössä. Myös mahdolliset MyData-mallit kannattaa rakentaa väestötietojärjestelmästä varmennettujen henkilötietoaineistojen päälle ja niihin pääsy tulisi rakentaa väestötietojärjestelmän sähköiseen asiointitunnukseen perustuvan vahvan sähköisen tunnistamisen varaan. Väestörekisterikeskus katsoo, että jos erilaisia MyData -malleja päätetään lähteä pilotoimaan, Väestörekisterikeskuksen tulisi olla mukana jo mallien kehitysvaiheessa.

Johtaja
Ylijohtajan sijaisena

Lea Krohns

Johtaja

Timo Salovaara

JAKELU LVM:n kirjaamo, Emil Asp/LVM

TIEDOKSI VRK kirjaamo, VRK:n johtoryhmä