

15.2.2016

Liikenne- ja viestintäministeriö
kirjaamo@lvm.fi

Viite: Lausuntopyyntö LVM/1996/13/2015

Lausunto luonnoksesta valtioneuvoston periaatepäätökseksi datan hyödyntämisestä liiketoiminnassa – massadan ja mydatan strategiset linjaukset ja toimet

Yleisiä havaintoja

Liikenne ja viestintäministeriön ehdotus käsittää useita ajankohtaisia ja keskeisiä huomioita sekä toimenpiteitä massadatan ja niin sanotun Mydatan tehokkaaksi hallinnoimiseksi ja yhteiskuntaa hyödyttäväksi soveltamiseksi liiketoiminnassa.

Tilastokeskus pitää hyvänä, että linjauksia valmisteltaessa on kiinnitetty erityistä huomiota tietosuojaan ja tiedon omistajuuteen, tietojen hyödyntämiseen päätöksenteossa sekä tietojen hyödyntämiseen liittyvään osaamiseen. Luonnos on kuitenkin laadittu yleisellä periaatetasolla ja useat sen toimenpide-ehdotukset edellyttävät jatkovalmistelua ennen yksityiskohtaisempaa kommentointia.

Vaikka periaatelinjaus on nimetty koskemaan datan hyödyntämistä liiketoiminnassa, linjauksissa käsitellään monin osin myös julkisen hallinnon roolia ja tehtäviä. Periaatelinjausten tavoitteiden ja kohdentumisen kirkastamiseksi voisi olla hyödyllistä tarkentaa dokumentin nimeä sekä siinä esitettyjen toimenpiteiden ja näiden vaikutusten kuvausta.

Dokumentista ei yksiselitteisesti ilmene, milloin puhutaan massadatasta, milloin datasta yleisemmin – tai toisaalta tarkoitetaanko datalla ns. yksikködataa vai laajemminkin eri tietovarantoihin syntyvää ja kertyvää tietoaainestoa (jalostettu tai aggregoitu data). Tämä vaikeuttaa joiltain osin linjausten ja toimenpiteiden toteuttamisen arviointia.

Kommentit periaatepäätösluonnoksen tavoitteisiin ja linjauksiin

Tilastokeskus on aktiivisesti mukana suunniteltaessa ja arvioitaessa massadatan hyödyntämistä tilastotoimessa niin kansallisella kuin kansainväliselläkin tasolla (YK, OECD, EU). Tilastokeskuksessa on käynnissä EU-rahoitteisia projekteja massadatan hyödyntämisessä esimerkiksi erilaisten indeksien muodostamisessa ja olemme mukana Euroopan tilastoviraston (Eurostat) massadatatyöryhmässä.

Tilastokeskuksen sisäinen Big Data –työryhmä laati tiekartan etenemisestä joulukuussa 2015. Erityinen paino on uusilla tietolähteillä, joista monet mahdollisuudet liittyvät liike-elämän omistamaan massadataan. Tavoitteena on toisaalta löytää keinoja tehostaa ja nopeuttaa tilastotuotantoa ja vähentää yritysten ja yksityishenkilöiden tiedonantorasitusta uusien tietolähteiden avulla, toisaalta löytää mahdollisuuksia uusien ilmiöiden kuvaamiseen ja uusiin tietotarpeisiin vastaamiseen.

Myös Tilastokeskuksen etäkäyttöpalvelulla (yksikköaineistojen käyttö tutkimustarkoituksiin) on kiinteä liittymä linjausehdotuksissa esitettyihin toimenpiteisiin.

Tilastokeskus esittää periaate- ja linjausluonnokseen seuraavat kommentit:

15.2.2016

Tavoite 2: Datan käytön kustannushyötysuhde kasvaa.

Linjauksessa todetaan mm. että ”Tavoitetilassa datan keruu ja hyödyntäminen ovat tehostuneet merkittävästi julkisen ja yksityisen sektorin yhteistyönä. Toimialoilla on käytössä yhtenäisiä tietomalleja ja –rakenteita, jotka mahdollistavat palvelukehitystä.”

Tilastokeskuksen näkökulmasta mahdollisuus yhdistää aikaisempaa laajemmin julkisen ja yksityisen sektorin aineistoja on tärkeää. Aineistojen yhdistely edellyttää yhteisiä tietomalleja ja –rakenteita, joihin olennaisesti liittyy tietojen sisällöllinen yhtenäisyys. Tähän tarvitaan yhteisiä tunnuksia, standardeja ja luokituksia (esimerkkeinä alueluokitukset, toimialaluokitus ja ammattiluokitus).

Periaatepäätöksessä olisi hyvä huomioida se, miten varmistetaan aineistojen luotettava yhdisteltävyys luokitusten, tunnusten ja muun sisällöllisen standardoinnin sekä metatietojen avulla.

Linjauksessa todetaan myös, ”uusia, tehokkaampia, julkisten ja yksityisten toimijoiden yhteistyöhön perustuvia keruumenetelmiä on mahdollistettava”.

Tilastokeskus pitää tavoitetta erittäin tärkeänä myös tilastotuotannon laadun ja ajantasaisuuden näkökulmasta ja on kiinnostunut osallistumaan soveltuvien osin kehitystyöhön.

Linjaus 1. Sääntely ja datapolitiikka

1.1 Tietoaineistojen laajempaa käyttöä edistetään kannustavalla sääntelyllä.

Linjaus on keskeinen tilastotoimen massadatan hyödyntämismahdollisuuksien kannalta. Ehdotukseen sisältyvä velvoite ”kriittisen datan” sekä julkisten tehtävien hoitoon liittyvän ja yleisen edun kannalta hyödyllisen datan jakamiseen ovat kiinnostavia.

Linjauksessa todetaan kokeilulakien käyttöönottoaminen mobiilidatan ja paikkatiedon hyödyntämiseen palvelukehityksessä viitaten liikenne- ja viestintäministeriöön. Tilastokeskus kuten kansainvälinen tilastotoimi ylipäättään on todennut teleoperaattoreiden mobiilipaikannusdatan erittäin keskeiseksi mahdollisuudeksi kehittää tilastointia.

Mobiilipaikannuksen keskeisimpiä käyttökohteita ja tarpeita tilastotuotannossa ovat matkailuun ja liikkumiseen liittyvä tilastointi, kansantalouden tilinpidon osana olevan matkailun tilinpidon laadun parantaminen sekä lakkautetun raja-haastattelututkimuksen tietojen korvaaminen matkustustaseen laskennassa.

*On tärkeää, että mobiilidatan käytön esteitä poistetaan myös virallisen tilastotoimen osalta ja aineistojen hyödyntämismahdollisuudet tilastotuo-
tantaan ovat mainittuina tässä linjauksessa.*

Linjauksessa ei mainita mitään datan säilyttämisestä. Mikäli tarkastellaan mahdollisuutta velvoittaa yritykset jakamaan yleishyödyllistä dataa, pitäisi linjauksessa ja toimenpiteissä ottaa kantaa myös datan säilytykseen, sillä se on palveluiden ja toiminnan jatkuvuuden kannalta olennaista.

1.2 Kyvykkyyttä kasvattavalla datapolitiikalla mahdollistetaan kasvu ja vahvistetaan päätöksenteon tietopohjaa.

Linjauksessa on todettu, että linjausten toteuttaminen edellyttää yhteistyötä useiden tahojen kanssa.

Tilastokeskuksen ja muun tilastoimen osallistuminen yhteisölliseen kehittämiseen on tärkeä huomioida jatkoa suunniteltaessa yhteiskunnallisten hyötyjen laajentamiseksi.

15.2.2016

Mikäli yhteiskuntaan halutaan uusia dataosaajia, on tärkeää, että opiskelijat ovat mahdollisimman aikaisessa vaiheessa tekemisissä oikean tai vähintään oikean muotoisen aineiston kanssa. Tilastokeskus on tuottanut ja voi jatkossakin tuottaa opetuskäyttöön soveltuvia aineistoja, joiden avulla julkishallinnon aineistojen yhdistämistä massadataan voidaan harjoitella.

Perustutkintojen kehittämisessä olisikin hyvä olla mukana tahoja, jotka käsittelevät ja joilla on pääsy yhteiskunnan keskeisiin tietoaineistoihin (esim. Tilastokeskus, THL ja Kela) kuin myös yksityisen sektorin datatointimijoita.

Linjauksessa todetaan, että tilastoviranomaisten kanssa on yhteistyössä kehitettävä datatalouden tilastointi. Tässä yhteydessä pitäisi muun muassa pystyä määrittämään ja tunnistamaan datatalouden toimialat.

Tilastokeskus pitää ehdotusta erinomaisena ja on valmis ottamaan keskeisen roolin asian edistämässä mikäli kehittämistyön resursointi on turvattu.

Linjaus 2. Datan palveluinfrastruktuuri ja käytänteiden kehittäminen

2.2 Kehitetään käytänteitä mahdollistamaan datan jakamista ja hyödyntämistä.

Keskeiseen rooliin linjauksen myötä nousevat tietosuojat, datan anonymisointi sekä eri tavat päästä käsittelemään tietoa (esim. etäkäyttö).

Tilastokeskuksessa on tehty pitkään työtä sekä datan anonymisoinnin että muiden tietosuoja koskevien käytänteiden kanssa.

Ennen tilastolain viimeisimmän muutoksen voimaantuloa 1.9.2013 (361/2013) Tilastokeskus muokkasi tutkimuskäyttöön tarkoitettuja aineistoja siten, että niistä ei ollut mahdollista välillisestikään tunnistaa yksilöitä. Lakimuutoksen jälkeen tilastoviranomaisilla on mahdollisuus antaa tutkimuskäyttöä varten käyttöoikeus (ei siis luovutus) aineistoon, josta ainoastaan suora tunnistaminen on estetty. Käytännössä tämä tarkoittaa sitä, että varsinaisen aineiston sijaan järjestetään Tilastokeskuksen etäkäyttöjärjestelmän kautta käyttöoikeus aineistoon, jossa tilastoyksiköille on muodostettu yksilöivä pseudotunniste. Tilastokeskus kehittää aineistojen pseudonymisointia edelleen siten, että eri lähteistä (esimerkiksi THL ja Tilastokeskus) tulevaa dataa voidaan teknisesti yhdistellä käyttäen pelkkiä pseudotunnisteita. Menetelmä kehitetään sellaiseksi, että se soveltuu myös massadatan ja valti-onhallinnon aineistojen yhdistelyyn.

Tilastokeskus kehittää yhdessä Kansallisarkiston kanssa kansallista rekisteri- ja mikroaineistojen tutkijapalvelua (Finnish Microdata Access Services FMAS). Palvelu on osa Suomen tutkimusinfrastruktuurien tiekarttaa. Rakennettavat palvelut parantavat merkittävästi julkishallinnon aineistojen löydettävyyttä, saatavuutta ja käytettävyyttä. Kokonaisuus koostuu keskenään yhteen toimivista palveluista, jotka tarjoavat yksikanavaisen väylän tutkimuksen suunnittelusta aineiston analysointiin. Tilastokeskus vastaa hankkeessa aineistojen metatietoluettelon ja etäkäyttöjärjestelmän toteuttamisesta, Kansallisarkiston vastuulla on käyttöluopien sähköisen hakupalvelun rakentaminen.

Etäkäyttö ja siihen liittyvät erilaiset anonymisointi-, pseudonymisointi- ja käyttöluparatkaisut tarjoavat oivan perustan laajennettaessa luvanvaraisten aineistojen käyttöä. Kehittämällä jo olemassa olevia ratkaisuja voitaisiin kohtuullisilla panoksilla helpottaa eri viranomaisten tunnuksellisten aineistojen yhdisteltävyyttä ilman varsinaisten tunnuksien vaihtamista.

15.2.2016

Tilastokeskuksen henkilöstölle on kertynyt vuosien mittaan merkittävää osaamista anonymisointiin, tilastotaulukoiden suojaukseen ja suojauksien suunnitteluun liittyen. Tilastokeskus on kiinnostunut olemaan mukana keskusteluissa, kun pohditaan anonymisointipalvelutoimijan perustamista.

Linjaus 2.3 Vaikutetaan valtion esimerkillä datan jakamisen synergiahyötyihin.

Tilastokeskus kannattaa kokeiluja ja uusien aineistojen pilotointia ajantasaisen, luotettavan tiedon tuottamiseksi päätöksenteon pohjaksi (evidence-based policy making). Tilastoinnin kehittämisessä massadata nähdään tässä keskeisenä mahdollisuutena.

Yhteyshenkilömme asiaan liittyen on kehittämisspäällikkö Pasi Piela (etunimi.sukunimi@stat.fi).



Marjo Bruun
pääjohtaja



Heli Mikkela
ylijohtaja

Jakelu

Liikenne- ja viestintäministeriö

TK/Kirjaamo, PJ, YJ, TYJ, Jussi Heino, Marja Sjöblom, Pasi Piela, Leena Storgårds