

Liikenne- ja viestintäministeriö
kirjaamo@lvm.fi
emil.asp@lvm.fi

TIEKEN lausunto kansallisen Big data –strategian luonnoksesta

TIEKE Tietoyhteiskunnan kehittämiskeskus ry kiittää mahdollisuudesta lausua näkemyksensä luonnoksesta kansalliseksi Big data -strategiaksi.

Strategiassa kiinnitetään hyvin huomiota big data haasteisiin, esimerkiksi tietoturvaan ja yksityisyyteen. Yhden keskeisen kysymyksen luonnos kuitenkin sivuuttaa, suurten tietomassojen automaattiseen analysointiin perustuvien johtopäätösten varmentamisen. Big datan tietomäärät ovat niin suuria, ja analyysin perustana oleva laskenta niin monimutkaista, että ihmisen on mahdollista tarkistaa, onko lopputulos oikea. Virheitä voivat aiheuttaa joko tahattomat virheet analysoitavissa tiedoissa tai analyysialgoritmeissa tai tietojen tai algoritmien tahallinen manipulointi. Kaikkia näitä vastaan on varauduttava, jotta luottamus tulosten oikeellisuuteen säilyy.

Lisäksi strategiassa tulisi keskustella monipuolisemmin sivulla 20 esitetystä yhteiskunnan luomasta luotetusta data operaattorista, joka operoisi analysointikoodia yritysten ja muiden organisaatioiden järjestelmissä. Ratkaisu olisi tiedonhallinnollisesti raskas, ja lisäksi merkittävä riski tietoturvalle ja yksityisyydelle.

Big datan mahdollisuuksien hyödyntäminen edellyttää lainsäädäntömuutoksia, ja niitä ohjaavaa informoitua julkista keskustelua esimerkiksi yksityisyyden suojasta, omien tietojen hallinnasta, tietojen omistajuudesta ja henkilötietolain periaatteista suhteessa big dataan. Esimerkiksi jos kansalainen omistaa omat tietonsa (my data), onko hänellä oikeus kieltää tietojensa big data -käyttö tai oikeus osuuteen tietojen käytön mahdollisista taloudellisista hyödyistä. Nykyinen lainsäädäntö myös estää joissakin tapauksissa henkilöön liittyvien tietojen yhdistämisen, vaikka henkilö antaisi siihen luvan. Esimerkiksi potilastietojen, verotietojen ja sosiaalihuollon tietojen yhdistäminen on sallittua vain manuaalisesti, ei automaattisesti.

TIEKE osallistuu mielellään big data –tietouden ja osaamisen levittämiseen ja julkisen keskustelun aktivointiin.

Ylipäättään tarvitaan julkista tietoisuutta big datasta, sen mahdollisuuksista ja haasteista, ei vain yrityksissä ja hallinnossa vaan myös kansalaisten keskuudessa. Strategialuonnoksen ”data-lähettiläiden” kohderyhmää tulisi laajentaa. Samoin, kuten luonnoksessa todetaan, tarvitaan avoin kokoava big data –klusteri luomaan jaettavaa tahtotilaa ja ratkaisemaan intressiristiriitoja.

TIEKE osallistuu mielellään Big data –klusterin toimintaan ja voi toimia myös ”Big Data Forum Finlandin” kokoajana. TIEKellä on kokemusta erilaisten klusterien ja foorumien kokoamisesta ja hallinnoimisesta. Esimerkkeinä Verkkomaksufoorumi ja ICT-klusteri.

Erityistä huomiota kannattaa kiinnittää Kanta-terveysarkiston tietojen hyödyntämisen pelisääntöihin. Vallitsevan käytännön mukaan potilastietoja ei saa käyttää tutkimukseen ilman henkilön suostumusta. Toisaalta Kanta-järjestelmä mahdollistaisi aivan uuden mittakaavan epidemiologista tutkimusta, jos tietojen käytölle löytyy toimivat tutkimuseettiset pelisäännöt.

Valitettavasti en korostaisi Suomen asemaa ”hyvin korkean tietosuojan maana”, etenkin IoT:n yhteydessä (sivu 6). Aalto Yliopistossa tehtyjen selvitysten mukaan suomalaisten Internetiin liitettyjen automaatiojärjestelmien tietosuoja on päinvastoin luvattoman heikko!¹ Vaikka Suomessa on maailmanluokan tietoturvaosaamista, se ei näy suomalaisten organisaatioiden toiminnassa.

Koulutus, esimerkiksi datatietoisuuden luominen sekä kansalaisille että organisaatioille on suuri haaste. Jo pelkästään opetussuunnitelmien päivittäminen ja opettajien täydennyskouluttaminen big datan ja muiden tietoyhteiskuntataitojen opettajiksi vaatii merkittäviä panostuksia.

TIEKellä on kokemusta sekä tietoyhteiskuntaosaamisen arvioinnista (TIEKEN tutkinnot ja tietotekniikkaosaamista arvioiva Tiviittori-palvelu) että tietoyhteiskuntaosaamisen täydennyskoulutuksesta. Big data –osaamisen ja datatietoisuuden lisääminen omiksi osioikseen TIEKEN tutkintoihin ja Tiviittori-palveluun² olisi suhteellisen helppoa, minkä jälkeen ne palvelisivat kaikkia big data –osaamista tarvitsevia organisaatioita.

Kuten luonnoksessa todetaan, big data edellyttää standardoituja tiedon esitysmuotoja ja sanomia, joilla tietoja siirretään tietojärjestelmien välillä. Standardit ja käytännöt ovat vasta muotoutumassa, ja Suomen tulisi olla tässä standardisointityössä aktiivisesti mukana. Toivottavasti kokeiluihin ja t&k-rahoitukseen kaavailtua rahoitusta osoitetaan myös tähän välttämättömään taustatyöhön.

TIEKE voi osallistua teknologiakartoitusten ja ”big data –etiketin” laatimiseen sekä mallien, määritelmien ja standardien tunnistamiseen ja kehittämiseen. TIEKellä on pitkä historia tieto- ja viestintäteknikan eri sovellusalojen kuten laskutuksen, logistiikan ja sähkönsiirron standardien kehittämisestä ja vakimuotoisten tiedonsiirtosanomien määrittelystä ja testaamisesta.

Erityisesti TIEKellä olisi annettavaa Älykkäät verkot (3.2) ja Liikenne (3.4) –sovellusalueilla. Molemmilla aloilla TIEKE on ollut mukana määrittelemässä nykyisiä käytäntöjä sekä tiedon siirtoon ja keräämiseen käytettyjä sanomia. Esimerkiksi nyt kun kuluttajista tulee älykkäissä sähköverkossa aktiivisia toimijoita, tarvitaan uusia sanomia ja käytäntöjä. Jatkokehitys kannattaa perustaa olemassa olevaan osaamiseen ja kokemukseen.

Vaikka tietosisällöt olisivat teknisesti yhteensopivia, myös tietosisältöjen määritelmien, metadatan on oltava yhteismitallisia, yhteentoimivia, muuten analyysien tuloksilla ei ole mitään merkitystä. Esimerkiksi Suomen sairaanhoitopiirit määrittelevät leikkauspäivän eri tavoin. Ennen käsitteistön yhtenäistämistä sairaanhoitopiirien keräämien tietomassojen yhdistäminen merkityksellistä analyysiä varten on mahdotonta. Käsitteiden määrittely ja sanastotyö ovat ehdoton edellytys big datan hyödyntämiselle, ja se tulisi huomioida strategiassa.

¹ <https://research.comnet.aalto.fi/public/Aalto-Shodan-Raportti-julkinen.pdf>

² Tiviittori-palvelu julkaistaan syksyllä

TIEKE on tehnyt jo pitkään sanastotyötä, jotta tietojärjestelmiin eri organisaatioissa ja organisaatioiden eri osassa kerätyt tiedot olisivat yhteismitallisia ja niitä voitaisiin hyödyntää täysipainoisesti. Esimerkiksi yhdessä THL:n kanssa on luotu terveydenhoitoalan sanaston määrittelemiseksi työväline, jonka avulla henkilöstö yhdessä osallistuu sanastotyöhön.

Julkisella sektorilla big datan jatkuvan analyysin hyödyntäminen toiminnan ohjaukseen edellyttää radikaalia valtion ja kuntien toimintakulttuurin muutosta, joka tulisi huomioida strategiassa. Nykyinen, vuosittain toistuvaan talousarvion ja strategian laadintaan perustuva ohjaus toimii huonosti nopeasti muuttuvassa maailmassa. Käytännössä loppuvuoden toiminta perustuu yli vuosi aiemmin tehtyihin arvauksiin.

TIEKE on mukana yritystietojen sähköistä viranomaisraportointia kehittävässä XBRL-hankkeessa, joka mahdollistaa mm. ajantasaiseen tietoon perustuvan talouspoliittisen ohjauksen (toissavuoden tilastojen asemesta) ja harmaan talouden ja veronkierron torjumisen. Paperilla kerättyjen tietojen varassa big datan mahdollisuudet eivät toteudu. TIEKE on mielellään mukana myös jatkotyössä, kun yritystietojen käyttöä laajennetaan nykyisestä (Verohallinto, Patentti- ja rekisterihallitus).

Kunnioittavasti

Jyrki Kasvi
Tutkimus- ja kehitysjohtaja
TIEKE Tietoyhteiskunnan kehittämiskeskus ry