

Liikenne- ja viestintäministeriölle

Viite: Lausuntopyyntö LVM138/05/2014, 3.6.2014

Big data -strategia

Strategia on selkeästi kirjoitettu ja antaa hyvän yleiskuvan. Sen vastapainona esitetyt toimenpiteet ovat kuitenkin kevyehkösti esitettyjä. Osittain toimenpide-ehdotukset jäävät tasolle jossa ehdotetaan vain että "asiaa edistetään", osa ehdotuksista taas tuntuu liittyvän kovin pieniin yksityiskohtiin.

Big data on määritelty laajasti, se tavallaan dokumentissa tarkoittaa teknologian kehittyessä yleistä siirtymistä uudenaikaiseen datan hallintaan, eli käsitettä "ei big dataa" ei ikään kuin ole. Sinänsä näkökulma on perusteltu, mutta jos ajatellaan että jokin data jää "big datan" ulkopuolelle, sellainenkin olisi hyvä määritellä, jotta strategiadokumentin käsitteellinen viesti saadaan selkeäksi.

Datan avaamista tukeva politiikka resurssitukineen on erinomaista, kunhan muistetaan että avoimen datan ylläpitäminen avoimessa muodossa sekin tarvitsee resursseja. Esimerkiksi valtionhallinnossa toimitaan supistuvilla budjeteilla, joissa avaamiseen investoiminen ei takaa avoimen datan tuotannon pysyvyyttä. Tämä asia on avoimesta datasta syntyvän yhteiskuntahyödyn kannalta kriittinen. Jos esim. yksityiset toimijat rakentavat liiketoimintansa julkisen sektorin avoimen datan päälle, on olemassa riski että julkisen sektorin leikkaukset vievät heiltä elinkeinon. Strategiassa voisi ottaa näkökulmaa siihen, miten turvataan julkisen sektorin toimintaa tietotuotannossa, joka tuottaa materiaalia kansantalouden kannalta tuottavasti yksityisen sektorin käyttöön.

Asianmukaisen yksityisyyden suojan huomioiminen tuodaan dokumentissa selvästi esiin. On hyvä, että tämän asian osaaminen voidaan ajatella asiaksi, jonka Suomi voisi kääntää vahvuudekseen.

Tutkimusta, siitä huolimatta että se big data -asioissa on keskeistä, käsitellään yllättävän pinnallisesti. Kohta 3.3. on vailla varsinaista sisältöä ja toimenpide-ehdotuksia koskevassa osiossa kohta 6.4. on tyhjä.

Kun big datan tutkimukseen ohjataan tutkimusrahoitusta, olisi tärkeää ohjata sitä myös ei-avoimen big datan käytön kehittämiseen. Tällä hetkellä on olemassa selvä riski siihen, että tutkimusrahoitus ohjautuu big data -hengessä avoimen datan ympärille, jolloin mm. sosiaali- ja terveydenhuollon tutkimus, jossa tutkimuskohde on usein yksilö, putoaa rahoitusmekanismien ulkopuolelle koska siellä ei säätelyn vuoksi ole helppoa muodostaa suuria dataa eri tavoin pyörittäviä tutkimusverkostoja.

Sivulla 22 olevaa Mydata-tekstiä olisi syytä vielä pohtia ennen strategian julkaisemista. Teksti on luettavissa tavalla jota ei voida pitää eettisesti hyväksyttävänä. On täysin eettisten periaatteiden vastaista ajatella, että yksityinen henkilö saisi parempaa yksilöllisempää ja halvempaa palvelua mikäli hän luovuttaa tietojansa yleiseen käyttöön. Tämä johtaisi siihen, että yhteiskunnan vähäosaiset, joilla ei ole edellytyksiä osallistua tällaiseen päätöksentekoon



16.6.2014

omissa asioissaan, saisivat huonompia palveluja. Lisäksi ajatus on tulkittavissa niin, että henkilöitä tavallaan kiristettäisiin luovuttamaan tietojaan jotta he saavat asianmukaisia palveluja. Todennäköisesti strategian kirjoittaja ei tarkoittanut tällaisia asetelmia, mutta näin teksti on nyt luettavissa.

Sivulla 23 oleva esimerkki terveydenhuollon osalta ei ole kovin onnistunut. Esimerkissä ei käsitellä dataa jota kertyy potilaiden hoitoprosesseissa, sillä tietokannoissa oleva data liittyy tieteellisen tutkimuksen kautta luotuun dataan. Päätöksenteon tukiohjelmat tietokannoista ovat Suomessakin jo pitkään olleet kliinisessä käytössä, joten esimerkki ei tuo esille mitään uutta. Suomi on yksi tämän osa-alueen edelläkävijämaista. Muutoinkaan terveydenhuoltoa koskevan osion 3.1. laadinnassa ei ole hyödynnetty alan asiantuntijoita. Vastuullisen aineiston käytön osalta eri osajien kanssa tehtävä yhteinen määrittely on täysin riittämätön toimenpide eikä ehdotuksella ole varsinaista sisältöä. Huomio siitä, että big datan hyödyntäminen nousee esiin sote-uudistuksessa, on perusteltu.

Kun terveydenhuoltoa kerran halutaan potentiaalisena osaamisalueena nostaa esiin, voisi kohdassa 3.5. huomata myös sairaalalogistiikan kuuluvan tälle osa-alueelle. Kohdassa 3.6. puolestaan ei käsitellä Cleantech ole määriteltä millään tavoin.

Terveyden sovellusalueella THL:n hallinnoimat tietovarannot mukaan lukien THL Biopankin varannot ovat keskiössä. THL on osaltaan ryhtynyt avaamaan dataansa suunnitelmansa mukaan. Biopankin osalta mietitään, missä muodossa sen tietosisältö (osittain dataa, osittain näytteitä joista on tuotettavissa lisää dataa) olisi parhaiten muiden hyödynnettävissä.

Tutkimuslaitoksen näkökulmasta big data avaa uusia näkymiä, joita ei toistaiseksi liene kovinkaan paljon tarkasteltu. Datalähtöinen toimintakulttuuri on tietovarantotyön ja avoimen datan myötä muotoutumassa, joskin kehitys on vasta alullaan. Big datan tarjoamat mahdollisuudet sote-palvelujentuottajien benchmarking-toiminnassa on avaus, jonka toteuttamiskelpoisuutta saattaisi olla hyödyllistä pohtia myös sote-rekisteridatan osalta. Pilottiprojektien tukeminen tässä voisi olla oikeansuuntainen aloite.

Terveyden sovellusalueella tehtävässä kehittämistyössä THL:llä voisi olla keskeinen rooli mittavien tietovarantojen hallinnoijana. Yhä supistuvien resurssien tilanteessa kuitenkin toimintaan on osoitettava erillistä rahoitusta.

Kun strategian toimeenpanosta tehdään suunnitelma, tulee siinä sopia vastuista ja resursseista.

Pääjohtaja

Juhani Eskola

Erikoissuunnittelija

Antti Tuomi-Nikula

www.thl.fi