



18.06.2014

Jakelussa mainituille

Viite LVM138/05/2014

Asia **OKM kansallisen big data strategian lausunto**

OKM on ollut mukana Big datan käyttö –työryhmässä laatimassa strategialuonnosta. Strategian tarkoituksena on tunnistaa ja löytää Suomen kannalta oleellisia alueita joilla big data voi tuoda arvoa niin yksityiselle kuin julkiselle sektorille. Big data on erittäin merkittävä asia myös opetus- ja kulttuuriministeriön toimialan kannalta.

Luonnoksen kappaleessa 1.1 esitellään big datan muutosvoimaa yritysmaailmassa ja julkisella sektorilla (esimerkkinä terveydenhuolto). Tähän on lisättävä että big data on tieteen alalla niin sanottu neljäs paradigma joka yhdistää kaikki kolme aikaisempaa paradigmaa eli kokeellisen, teoreettisen ja laskennallisen. Yhdessä avoimen datan kanssa se mullistaa myös tiede- ja tutkimusmaailman.

Samassa kappaleessa sanotaan että: ”Ennen valtavan ja erimuotoisen tietomäärän osittainen tallentaminen, yhdistely ja analysointi ei ollut mahdollista.” Vaikka analysointi voi käsittää monenlaisia asioita niin tässä on syytä erikseen mainita että valtavien tietomäärien visualisointi on ketjun seuraava vaihe ja iso haaste sekkin.

Kappaleessa 1.2 puhutaan tiedon tulkinnan merkityksestä. Tässä olisi paikallaan myös pohtia datan kompleksisuutta, tai ainakin tuoda esille se että dataintensiivisen laskennan suurimpia ongelmia on datan strukturoimattomuus.

Kappaleessa 1.3 tulee tarkentaa että big data ja avoin data ovat osa Euroopan Komission Digital agenda for Europe -kärkihanketta. Siihen liittyvä puiteohjelma on nimeltään Horizon 2020, ja siitä rahoitetaan eurooppalaisia tutkimus- ja innovointihankkeita vuosina 2014–2020 yhteensä lähes 80 miljardilla eurolla. Komission tavoitteena on luoda tämän puiteohjelman avulla kasvua ja uusia työpaikkoja sekä parantaa yritysten asemaa globaalissa kilpailussa.

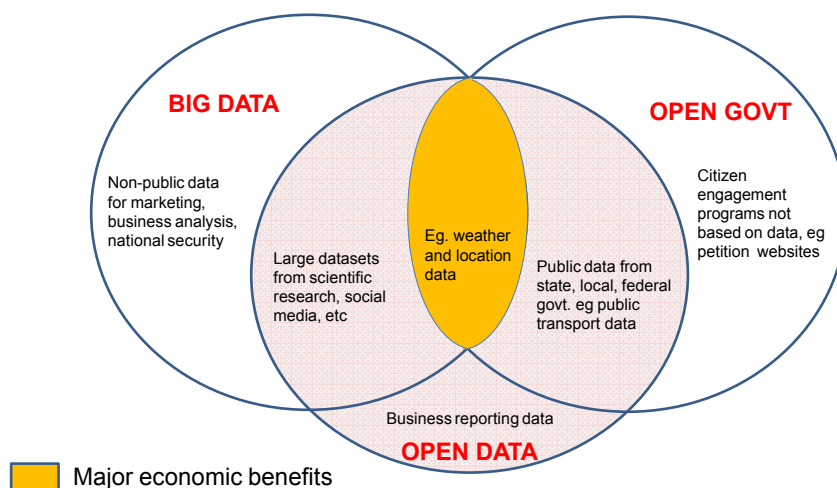
Kappaleessa 1.4 mainitaan Suomen hyvät edellytykset big datan hyödyntämiseen. Listaa tulisi jatkaa mainitsemalla suomalaista opetusta ja tutkimusta palveleva huippunopea ja luotettava Funet-tietoverkko.

Kappaleessa 2.4 tai kappaleessa 2.8 (tai molemmissa) tulee mainita että tekijänoikeudelliset epäselvyydet toistaiseksi estävät Suomessa tekstin- ja datanlouhinnan (TDM) jolla on etenkin humanistisilla aloilla kuten kielitieteissä ja kirjallisuudessa vahvasti tutkimusta haittaava merkitys. USA:ssa sama asia on ratkaistu niin sanotulla Fair Use -säännöksellä joka sallii TDM:n tutkimuskäytön.

Kappaleessa 2.6 tulee erikseen mainita koneluettavan metadatan tärkeys. Tieteenalakohtaisilla metadatatamalleilla kuvaillaan tietoaaineistot ja mahdollistetaan tietoaaineistojen oikea jatkokäyttö. Joillekin tieteenaloille kuten yhteiskuntatieteille on jo olemassa metadata standardit, osalle niitä ollaan vasta kehittämässä.

Kappaleen 2.7 kuva tulee piirtää uusiksi. OKM:n kalvopohja on turha siinä yhteydessä. Parempi on käyttää alkuperäistä lähdettä <http://www.theguardian.com/public-leaders-network/2014/apr/15/big-data-open-data-transform-government> tai alla olevaa uutta hahmotelmaa kuvasta. Hyödyntämisen kannalta mielenkiintoisimpia alueita on big datan ja julkishallinnon avoimen datan risteyskohta.

Data with the most economic benefits



Kappaleessa 2.9 puhutaan yhteistyöstä ja datan vaihdosta. Tieteessä ja tutkimuksessa datan jakaminen ei vielä olen kovin yleistä. Sitä auttaisi jakamisesta annettava selkeä meriitti, samaan tapaan kuin tieteellisistä julkaisuista. Muutamat rahoittajat jo vaativat datanhallinnan suunnitelmaa (DMP) mutta datan avoimeksi saattaminen on vielä suosituksien varassa. Datan määrän kasvaessa saattaa tulla vastaan tilanne jossa julkisella puolella joudutaan valitsemaan pitempään säilytettävät aineistot. Valintaan vaikuttavia kriteerejä voivat olla esimerkiksi aineistojen kypsyys ja arvo.

Kappaleessa 3.1 ehdotamme mainittavan eurooppalaisiin tutkimusinfrastruktuureihin kuuluvan ELIXIR –hankkeen joka on keskeinen Suomen bio- ja lääketieteellisen tutkimuksen kannalta.

Kappaleessa 4.2 on aiheena kansalaisille yksilöllisempää julkista palvelua. Yhtenä kantavana ajatuksena on kansalaisten segmentointi. Tässä voisi nostaa esiin erityisesti mahdollisuudet nuorten parissa tehtävässä työssä, niin liikunnan- kuin nuorisotyön osalta. Kappaleessa todetaan että Big

Datan käyttö on keskeinen keino parantaa vaikuttavuutta ja kansalaistyytyväisyyttä tarjoamalla henkilökohtaisempaa palvelua. Tämän tyyppiset sovellukset olisivat erittäin tervetulleita liikuntatottumusten vahvistamisessa vähän liikkuville sekä esimerkiksi syrjäytymisvaarassa olevien nuorten sosiaalisessa vahvistamisessa. Järjestöjen vetovoima on heikentynyt, on tarvetta löytää julkisia palveluja jotka tukevat nuorten kiinnittymistä yhteiskuntaan sekä tukevat nuoria esimerkiksi liikunnalliseen elämäntyyliin on etsittävä järjestömaailman ulkopuolelta.

Kappaleessa 5.3 on tuotu esiin joukkoistaminen ja sen mahdollisuuksia parantaa demokraattista osallistumista sekä kansalaisten äänen kuulumista päätöksenteossa. Nuorten parissa mobileteknologian käyttö on korkeata, ja mahdollisuudet lisätä kiinnostusta demokraattista päätöksentekoa kohtaan perinteisen puoluepoliittisen kanavan lisäksi ovat isot. Vaikka ei sinänsä ole suora yhteys Big datan ja sähköisten äänestysjärjestelmien välillä niin tässä voitaisiin vahvemmin tuoda esiin mahdollisuudet.

Kappaleessa 5.4 on aiheena tiedon etsintä. Tähän liittyy oleellisesti semanttinen verkko ja ontologiat teknologisenä ratkaisuna. Suomessa on lähdetty jo 2003 kehittämään semanttisen webin ideologian mukaisia teknologisia ratkaisuja. Parhaillaan Kansalliskirjastossa on menossa ONKI-projekti jossa kiinnitetään huomiota ontologioiden käytettävyyteen ja sovellettavuuteen erilaisissa yhteyksissä.

Kappaleessa 5.5 käsitellään paikkatietoa. Kappaleessa todetaan että paikkatietojen innovatiivinen käyttäminen tarjoaa aivan uusia näköaloja kuljetuspalveluiden, yhdyskuntasuunnittelun ja liiketoiminnan suunnittelun suhteen. Palveluverkostojen optimaalinen kohtaavuus palveluja haluavan kanssa voidaan suunnitella täysin vedenpitävästi. Myös tässä kappaleessa voisi esimerkiksi mainita nuorille suunnatut palvelut yleensä, mutta erityisesti palvelujen yhteensovittaminen syrjäytymisvaarassa oleville nuorille, selkeät palveluverkostot ja -ketjut ovat oleellinen osa syrjäytymisen ehkäisemistä ja keskeinen tavoite monialaisessa yhteistyössä.

Kappaleessa 6.2 on esitetty kuinka opetushallituksen johdolla muotoiltaisiin tutkintoihin big datan vähimmäistaidot. OKM toteaa ettei edellä mainittu ole OPH:n rooli lainkaan, eikä nykyainsäädännön mukaan toivottava kehityssuunta.

Esitetyt toimenpiteet ovat kannatettavia. Seuraavassa OKM:n kannalta relevantteja tarkennuksia.

Koulutukseen (6.2) ja osaamiseen (6.3) liittyvät ehdotukset (d-g sekä h-k) ovat tärkeitä, mutta edellyttävät uudelleen muotoilua, kohdentamista tai poistamista. Toteamme, että d) Perusopetuksen OPS perusteet valmistuvat vuoden 2014 loppuun mennessä. Sitä edeltävä virke ”EU:n tavoitteena on ...” ei liity mitenkään OPS perustetyöhön. Esitämme myös harkittavaksi, onko esitetty ehdotus relevantti tässä yhteydessä.

Myös opettajankoulutukselle ehdotetut toimenpiteet (e-g) tulisi muotoilla uudestaan (niiltä osin kuin koskevat opettajankoulutusta antavia korkeakouluja). Termit, kuten ”big data opetustaidot” tulisi avata. Lisäksi tulisi varmistaa se, että ehdotettu toimenpide on ehdotetulla tavalla kartoitettavissa, koordinoitavissa tai osallistuttavissa.

Soveltavan osaamisen osalta tulisi tarkentaa ehdotukset keskeisiksi toimenpiteiksi alan osaajien ja osaamisen varmistamiseksi. Tässä kannattaa ottaa huomioon ICT2015 –toimenpideohjelman Big Dataa koskevat ehdotukset.

Tärkeitä big datan käytön kannalta ovat myös sääntelyyn liittyvät kohdat (t-v). OKM ehdottaa, että samassa yhteydessä pohdittaisiin myös tekijänoikeuksiin liittyviä kysymyksiä.

Kappale 6.4 käsittelee tutkimusta. Ymmärryksemme mukaan korkeakouluja on pyydetty itse tuomaan esille omat aiheeseen liittyvät tutkimusfokukset. OKM:llä ei tätä tietoa ole olemassa ja toivon mukaan tästä saadaan arvokas tietopaketti big datan tutkimuksesta. Tässä kappaleessa on syytä nostaa esille OKM:n CSC:ltä ostamat korkeakouluille suunnatut palvelut liittyen dataintensiiviseen laskentaan. Yhtä lailla on tärkeä mainita avoimia tietoaaineistoja sivuava Avoin Tiede ja Tutkimus – hanke jossa on tarkoitus luoda Suomeen maailman huippua olevat tutkimusedellytykset

Kansliapäällikkö

Anita Lehikoinen

Ylitarkastaja

Sami Niinimäki

Jakelu Liikenne- ja viestintäministeriö
Suunnittelija Emil Asp

Tiedoksi Ylijohtaja Eeva-Riitta Pirhonen
Ylijohtaja Tapio Kosunen
Ylijohtaja Riitta Kaivosoja
Ylijohtaja Rauno Anttila
Johtaja Kirsi Kangaspunta