

Viite: LVM/1996/13/2015

Asia: Lausunto luonnoksesta valtioneuvoston periaatepäätökseksi datan hyödyntämisestä liiketoiminnassa – massadatan ja mydatan strategiset linjaukset ja toimet

Lausunnon antajan tiedot:

Organisaatio:	OKM
Lausunnon laatijat:	Juha Haataja, Sami Niinimäki

Yleiset kommentit

- Nyt lausuntokierroksella oleva luonnos periaatepäätökseksi vastaa sisällöltään aiempaa vuoden 2011 periaatepäätöstä, jossa on mukana mm. liike-elämän ja innovaatiotoiminnan tarpeet ja jossa on tunnistettu osajatarve. LVM:n valmistelema valtioneuvoston periaatepäätös julkisen sektorin digitaalisten tietoaineistojen saatavuuden parantamisesta ja uudelleenkäytön edistämisestä julkistettiin 3.3.2011 (<http://valtioneuvosto.fi/paatokset/periaatepaatokset>). Uudessa periaatepäätöksessä on syytä huomioida erityisesti alla mainitut seikat.
- Tekstissä tulee selkeyttää, mikä on periaatepäätöksen asema nykyisen tietohallintolain kontekstissa, kun ehdotettujen linjausten pääsisältö asettuu toimenpiteisiin sekä rakenteiden ja käytäntöjen luomiseen. Tämän vuoksi
 - o Ehdotetusta periaatepäätöksestä pitää erottaa tietohallintolain ohjaukseen kuuluvat asiat ja täsmentää periaatepäätöksen fokus yleisemmälle tasolle.
 - o Periaatepäätösehdotus sisältää huomioita kuten ”linjaukset tukevat tietojen hyödyntämistä luottamusta herättävällä tavalla”, mikä käytännössä liittyy sekä tietosuoja-asetuksen toimeenpanoon että tietohallintolakiin. Näiden yhteensopivuus ehdotettujen linjausten kanssa on syytä tarkastaa.
- Käsitteiden määrittelyt tulee tehdä huolellisesti ja täsmällisesti. Esim. data-termiä käytetään eri konteksteissa ja merkityksissä. Muun muassa seuraavia asioita pitää täsmentää:
 - o Data liittyy tasoon, jossa etsitään korrelaatioita symbolien, mittaustulosten ja metriikan sisällöstä.
 - o Informaatio liittyy tasoon, jossa on jo korreloimalla tunnistettu hahmoja, tehty vertailuja ja järjestetty dataa. Analytiikka kohdistuu tähän.
 - o Analyysin tuloksena saadaan trendejä, yleistyksiä jne., joita voidaan soveltaa.
 - o Luonnoksessa sekoitetaan digitalisaatio, massadata sekä tietopohjainen toiminta terminologisesti ja sisällöllisesti.

- "Data mahdollistaa digitalisaation hyödyntämisen..." pitäisi olla pikemminkin "Digitalisaatio mahdollistaa datan hyödyntämisen..."
- Luonnoksessa lähdetään instrumentaaliselta pohjalta ja ajatellaan että tietojen hyödyntäminen koostuu teknisistä toimenpiteistä ja palvelutarjoomasta. Osassa tapauksista näin onkin, mutta todella radikaalit ja merkittävät liiketoiminnan kasvattamisen mahdollisuudet (jos analytiikka tuottaa tuloksia) ovat nimenomaan sisällöllisellä puolella, analyysien tuloksissa ja niiden pohjalta syntyvissä tuoteideoissa. Suomea moititaan jo alhaisen digitalisaation maaksi, ja ehdotuksen pitäminen tällä tasolla vahvistaisi varsin matalan tason toimintaa, ts. massadatan rajapintaan kiinnittyvien tietoteknisten palveluiden rakentumista. Olennaista edelläkävijyydessä on ymmärtää toiminnan substanssi eikä lähteä välinekeskeisestä lähtökohdasta.
- Erityisen tärkeää on saada yrityksiä mukaan yhteiskehittämiseen. Näyttää että luonnoksen laatimiseen on osallistunut vain yksi yritys (OKM:n erityistehtäväyhtiö CSC – Tieteen tietotekniikan keskus Oy), jolle on laajasti ehdotettu roolia, myös tehtävissä jotka kuuluvat lähtökohtaisesti ministeriöille ja näiden alaisille virastoille. Tämä näyttäytyy valtioneuvoston periaatepäätöksen kontekstissa problemaattiselta koko yrityskentän näkökulmasta, ja voi tuntua yhden yrityksen piilotuolta.
- Osaamisen kehittyminen ja sen hyödyntäminen on tunnistettu Suomen innovaatiojärjestelmän kehittymisen avainkysymykseksi monien vuosien ajan monen toimijan suulla. Nykyiset koulutukseen ja tutkimukseen kohdistuvat leikkaukset vähentävät osaamista. Mahdolliset toimenpiteet tulee kohdistaa OKM:n kautta osaamisen syntyyn ja hyödyntämiseen.
- Luonnoksessa puhutaan Suomesta houkuttelevana datapalveluiden tarjoajana ja kumppanina. Tässä varmaankin tarkoitetaan suomalaisia liikeyrityksiä. Luonnoksessa tuntuu sekoittuvan julkishallinnon rooli sekä liiketoiminta. Esimerkiksi luonnoksessa ehdotetun avoimen palveluinfrastruktuurin kustannukset on syytä saada liikevoiton saavien tahojen maksettavaksi.
- Tutkimustieto on ennen kaikkea osaamista, joka ilmenee syntyvinä tutkimustuotoksina. Tärkeimmät tutkimustuokset ovat julkaistavat artikkelit, opinnäytteet ja selvitykset. Tutkimusta tukemaan kertyy ja syntyy myös erilaisia muita tutkimusmateriaaleja, kuten havaintotietoja. Osaamisen ja julkaistujen tutkimustulosten hyödyntäminen liiketoiminnassa on yritysten t&k-toiminnan ydin. Luonnoksen toimenpiteiden tulee kohdistua tähän. Toimenpiteeksi sopii parantaa yritysten mahdollisuuksia saada käyttöönsä tutkimusjulkaisuja.
- Luonnoksessa puhutaan useassa kohdassa tiedon siirtämisestä. Kuitenkin kansainväliset yritykset ovat raporteissaan ja analyyseissään päätyneet tulokseen että tietovarantoja on parasta säilyttää ja hyödyntää primäärilähteestään modernin hajautetun SOA-arkkitehtuurin pohjalta. Toisaalta luonnos sisältää ehdotuksen myös tästä. Olisi tarpeen, että luonnos täsmentää ja visualisoi tavoitetilaansa ja sitä mihin toimenpiteet tulee kohdentaa.

- Datatiede on leikkaus tilastotiedettä, tietojenkäsittelytiedettä ja substanssialaa. Massadata, jota periaatepäätös pitkälti käsittelee, on datatieteen rajattu osa-alue, vaikka ehdotuksessa niitä synonyymeinä käytetäänkin. Ottaen huomioon periaatepäätöksen kontekstin on tarpeen, että käytetään termiä jolla on laajempi merkitys kuin massadatalla.

Tiivistelmä

- Luonnoksessa puhutaan ”laajaa hyötyä edustavista tietoaaineistoista”. Olisi hyvä tunnistaa näitä konkreettisesti, ja varsinkin yrityksiä jotka näiden pohjalta voisivat rakentaa liiketoimintaa.

Strategiset tavoitteet

- Konkretiaa ja selvennystä kaivataan kohtaan jossa visioidaan: *”Valtiolla voi olla esimerkiksi omistajan rooli markkinoiden avaamisen alkuvaiheessa. Toisaalta julkinen sektori voi myös luopua palvelujen tarjoamisesta, jossa tehtäviä voidaan dataperusteisesti hoitaa markkinaperusteisesti”*
- Luonnos sisältää yleisiä huomioita kaikkien osapuolten hyödystä, mutta näitä ei kuitenkaan konkretisoida. Yleiskuvaksi jää että julkishallinnon tulisi vähenevien resurssien paineessa ohjata toimialojen kehittämisen rahoitusta liiketoiminnan edellytysten tukemiseen. Sen sijaan olisi parempi, että ehdotettujen liike-elämän ja korkeakoulujen yhteisten infrastruktuurien päärahoittajana olisi liike-elämä, koska rahallinen hyöty kertyy yrityksille. Julkishallinnon panos olisi esim. viranomaistiedon tuominen infrastruktuuriin.

Tavoite 1: Datalla ja analytiikalla älykkäämpää liiketoimintaa

Tavoite 2: Danan käytön kustannushyötysuhde kasvaa

- Koska ei ole osoitettu rahallisia hyötyjä joita julkishallinnon panos yritysten tukemiseen infrastruktuurilla tuottaisi, on vaikea osoittaa kustannushyötysuhteen kasvua. Tähän on hyvä lisätä konkretiaa.

Tavoite 3: Palvelutarjonta kansainvälistyy

- Tässä on hyvä olla enemmän pohdintaa siitä, miten suomalaista kilpailukykyä voidaan tukea. Onko kipukynnys puuttuva infrastruktuuri vai tuotteiden paketointi, informaatiomuotoilu ja markkinointi?

Tavoite 4: Kyvykkyys ihmiskeskeiseen datan käyttöön ja henkilötietojen hallintaan nousevat Suomen kilpailukykytekijäksi

Taustaa

- Korkeakoulujen ICT-osaamiskompetenssien osalta voisi viitata OKM-ICT-2015-hankeryhmän tekemään korkeakoulujen ICT-profilointikyselyyn, josta löytyy tietoja netistä:

<http://www.okm.fi/OPM/Tapahtumakalenteri/2014/10/OKM ICT 2015 seminaari.html?lang=fi>

- Luonnoksessa mainittu avoimen tieteen ja tutkimuksen tiekartta pyrkii toimenpiteillään edistämään avoimia toimintamalleja tieteellisessä tutkimuksessa niin että tutkimustieto liikkuu sujuvasti koko yhteiskunnassa.

Taloudellinen merkitys

Valtioneuvoston linjaukset ja toimenpiteet massadatan hyödyntämiseksi liiketoiminnassa

1. Sääntely ja datapolitiikka

Linjaus 1.1 Tietoaineistojen laajempaa käyttöä edistetään kannustavalla sääntelyllä

- Luonnoksessa todetaan että tietopohjaisiin liiketoimintamalleihin siirtyminen tuo huomattavaa kilpailuetua yrityksille, toisaalta todetaan että yrityksille pitää saada tähän lisätukea. Jos kilpailuetu on ilmiselvä, kuuluu yritysten normaaliin tuotekehitykseen ja investointikohteisiin lähteä kehittämään aluetta. Esimerkiksi SLUSH-tapahtumassa näkyi kiitettävästi uusia liiketoimintaideoita, joille haettiin rahoitusta. Ehdottaako luonnos että julkishallinto alkaisi kasvattaa rooliaan tässä rahoituksessa?

”Arvioidaan dataliiketoiminnan kytköksiä tekijänoikeuslainsäädännön kehittämiseksi esimerkiksi sitä, miten ja missä määrin tietoaineistojen vapaata käyttöä (public domain) sekä teksti- ja datalouhintaa voidaan edistää Euroopan sisämarkkinatoimien mukaisesti. (OKM)”

- Teksti tulisi kirjoittaa seuraavaan tyyliin: ”Arvioidaan tarpeet tekijänoikeuslainsäädännön kehittämiseksi, esimerkiksi laatimalla aiemmin tehtyjen selvitysten pohjalta etenemissuunnitelma (OKM).”

Linjaus 1.2 Kyvykkyyttä kasvattavalla datapolitiikalla mahdollistetaan kasvu ja vahvistetaan päätöksenteon tietopohjaa

- Kyvykkyuden kasvattamisessa on mainittu toimijana CSC – Tieteen tietotekniikan keskus Oy kohdissa, jossa k.o. yrityksellä ei ole toimivaltaa, sillä toimijana on esimerkiksi OKM ja korkeakoulut. On syytä poistaa tällaiset maininnat.

”Lisätään panostusta opetukseen ja huippututkimukseen aihepiiriä tukevissa tieteenaloissa. Kansallisen osaamis- ja tutkimuspääoman kasvattamiseksi on koulutusta ja tutkimusta ripeästi kehitettävä ja kohdennettava elinkeinoelämän tarpeisiin (OKM/CSC).”

- Tekstiehdotus: ”Korkeakoulut lisäävät panostusta opetukseen ja huippututkimukseen aihepiiriä tukevissa tieteenaloissa. Kansallisen osaamis- ja tutkimuspääoman kasvattamiseksi on varmistettava koulutusta ja tutkimusta ripeästi kehittämällä ja elinkeinoelämän tarpeisiin kohdentamalla (OKM/korkeakoulut).”

- Datan hallinnan ja analytiikan koulutusta tehdään varsin monipuolisesti korkeakouluissa. Olisikin hyvä jos luonnoksessa ehdotettu koulutuksen vahvistamista avattaisiin ja selkeytettäisiin, eli tarkoittaako tämä koulutustarjoonnan (eli erilaisten kurssien) tarvetta (ja jos tarkoittaa, niin mitä koulutusta tulisi lisätä) vai liike-elämän toimijoille paremman osaamisen saamista (mikä on järjestettävissä esim. osaavien henkilöiden palkkaamisella)? Vai tarkoitetaanko tässä esimerkiksi korkeakoulujen koulutuksen kehittämistä?

”Tunnistetaan elinkeinoelämän osaamistarpeita ja tarvetta alan osaamiskeskittymän organisoimiselle (LVM/CSC)”

- Osaamiskeskittymän hahmottaminen kuuluu ministeriöille (OKM ja TEM) sekä korkeakouluille ja tutkimuslaitoksille. OKM:n erityistehtäväyhtiö CSC – Tieteen tietotekniikan keskus Oy ei voi olla edustamassa koko yrityskenttää. Tekstiehdotus: *”Tunnistetaan elinkeinoelämän osaamistarpeita ja tarvetta alan osaamiskeskittymän organisoimiselle (TEM, OKM, LVM ja muut ministeriöt).”*

”Oppilaitosten yhteistyölle koulutuksessa ja tutkimuksessa luodaan kannusteita. Luodaan datatieteilijöille huippukyvykkyyksien kehityspolut ja kehitetään perustutkintoihin dataosaamisen osakokonaisuuksina eri tieteenaloilla. Dataosaajien yhteisöjen kehittämisellä tuetaan liiketoiminnan ja opetuksen tarpeiden sovittamista. Kouluttajien kouluttamista on lisättävä ammattikorkeakoulutasolla (OKM).”

- Yllä olevaan tekstiin liittyen on huomioitava seuraavaa:
 - o Koulutuksen sisältöjä on uudistettava huomioiden työelämän tarpeet ja työelämän muutokset, ja osaamista on hyödynnettävä nykyistä laajemmin. Elinkeinoelämän kanssa on tehtävä kumpaakin sektoria hyödyttävää yhteistyötä.
 - o Osaamisen tulee syntyä liiketoimintaan. Jos pullonkaulana on koulutuksen ja liike-elämän tarpeiden kohtaamattomuus, on syytä vahvistaa korkeakoulujen ja liike-elämän osaajatuotannon sekä tarpeiden kohtaamista. Tämän pohjaksi voi selvittää massadatan ja tietopohjaisen toiminnan osaajien työllistymistä Suomessa sekä luoda rekrytointikanava ja yhteisprojektien mahdollisuudet korkeakouluihin.
 - o Toimenpiteiden osalta olisi etua, jos yleisen liike-elämän kehittymisen sijaan olisi tunnistettu toimialat, joita voitaisiin ensimmäisinä alkaa tukea tässä toiminnassa. Suomen painopistealueet (clean tech, biotalous etc.) olisivat luontaisia lähtökohtia, joissa olisi mietittävä tietopohjaisuuden tuoman liiketoimintaedun konkretisointia.

”Tutkimusinfrastruktuurien tuottaman datan hyödyntämiseen ja tutkimusinfrastruktuurien yhteiskäyttöön sekä tutkimuksen ja sen tuottamien aineistojen, tulosten, osaamisen ja menetelmien muuntamiseen liiketoiminnaksi on panostettava toimintamalleja kehittämällä ja tiivistämällä osaamisen siirtoa ja yhteistyötä korkeakoulujen, tutkimuslaitosten ja yritysten välillä (OKM, TEM, LVM).”

- Teksti tulisi jakaa kahteen osaan ja muuttaa seuraavaan muotoon:

- ”Tutkimusinfrastruktuurien tuottaman datan hyödyntämistä ja tutkimusinfrastruktuurien yhteiskäyttöä vahvistetaan (OKM, TEM, Tekes, Suomen Akatemia).”
- ”Tutkimuksen ja sen tuottamien aineistojen, tulosten, osaamisen ja menetelmien laaja-alaiseen hyödyntämiseen sekä muuntamiseen liiketoiminnaksi on panostettava toimintamalleja kehittämällä ja tiivistämällä osaamisen siirtoa ja yhteistyötä korkeakoulujen, tutkimuslaitosten ja yritysten välillä (OKM, TEM, LVM).”

”Järjestetään suurten tietoaaineistojen hyödyntämiseen ja analytiikkaan liittyvää täydennyskoulutusta (OKM).”

- Ehdotus tekstin muotoiluksi: ”Huolehditaan tarvittavan osaamisen saatavuudesta yritysten käyttöön (OKM).”

”Kannustetaan tutkimuslaitosten, korkeakoulujen, kansalaisten (kansalaistiede) sekä yritysten osallistumista yhteisölliseen työhön, jonka tuloksena syntyy jaettu tietoaineistoja.”

- Ehdotus tekstin muotoiluksi: ”Kannustetaan tutkimuslaitosten, korkeakoulujen, kansalaisten (kansalaistiede) sekä yritysten osallistumista yhteisölliseen työhön, jonka tuloksena syntyy uutta tietoa tai liiketoimintaa.”

2. Datan palveluinfrastruktuurin ja käytänteiden kehittäminen

- Luonnoksessa ehdotetaan kaikille yhteisten käytäntöjen, menettelyiden ja mallien kehittämistä. Epäselväksi jää, mistä yritykset jatkossa voisivat erottua toisistaan ja saada kilpailuetua. Monen kasvuyrityksen lähtökohta perustuu sosiaaliseen innovaatioon, jossa tietoa hyödynnetään uudella tavalla. Liian tiukka yhteisten menettelyiden ja mallien vaatiminen katkaisee yritteliäisyyden mahdollisuuksia.

Linjaus 2.1 Tuetaan datan hyödyntämiseen perustuvien toimintamallien kehittämistä ja markkinoiden avautumista

Linjaus 2.2 Kehitetään käytänteitä mahdollistamaan datan jakamista ja hyödyntämistä

- Yhteentoimivuuden ja standardoinnin etenemiseksi ei riitä tiedon jakaminen toimialakohtaisista de facto-käytänteiden kehitystyöstä ja esimerkeistä vaan tulee aktiivisesti panostaa kansalliseen ja kansainväliseen yhteistyöhön ja edistää toimijoiden yhteistyötä esimerkiksi EU-tasolla (esim. Public-Private Partnership -lähtökohdasta).

Linjaus 2.3 Vaikutetaan valtion esimerkillä datan jakamisen synergiahyötyihin

3. Dataa hyödyntävän analytiikka-, palvelu- ja sovelluskehityksen edistäminen

- Luonnoksessa ehdotetaan keskeisten kansallisten toimijoiden rakentamaa ekosysteemiä. Jos nykyisten julkishallinnon palveluiden tuottamista yrityksien avulla halutaan edistää,

miksi valtion tulisi rakentaa tähän ekosysteemi ja kattaa sen kustannukset? Eikö olisi syytä antaa yritysten rakentaa ekosysteemi?

- Tällä hetkellä käytössä on JulkICT-lab, joka on perustettu kyseessä olevan ekosysteemin siemeneksi. Ennen kuin konseptia laajennetaan, olisi syytä tuoda ilmi saadut kokemukset ja selvittää, millaista hyötyä tähänastinen toiminta on tuonut.

Linjaus 3.1 Datan soveltaminen yhteiskehittelyn ja yhteistyön kautta

- Luonnoksessa esitetty keskitetty informaatiohubi kuulostaa vanhakantaiselta ajattelulta. Enemminkin tulee luoda mahdollisuuksia hajautetulle toiminnalle. Kaikkien munien paneminen yhteen koriin varsinkin liike-elämässä tuntuu jäykältä tavalta toimia.
- Tutkimustiedon kokoaminen kansalliseen informaatiohubiin ja siirtäminen ESFRI-infrastruktuureista ulos tarkoittaisi tiedon hyödyntämiseen tarvittavan osaamisen irtikykentää. Samoin liike-elämällekin palveluita tarjoavat tutkimusinfrastruktuurit menettäisivät tulolähteensä mikä vähentäisi niiden mahdollisuuksia toiminnan kehittämiseen.
- Nykyiset datankäsittelyratkaisut perustuvat pilvipalveluihin, jotka fyysisesti saattavat sijaita missä päin maailmaa tahansa. Erityistä huomiota tulee kiinnittää tietoturva-asioihin ja varmistaa käsiteltävien datojen sensitiivisyysaste ja kohdemaan tietosuojan taso.
- Lähtökohtana tulee olla että kannustetaan tutkimuslaitosten, korkeakoulujen, kansalaisten (kansalaistiede) sekä yritysten osallistumista yhteisölliseen työhön, jonka tuloksena syntyy uutta tietoa tai liiketoimintaa.

Linjaus 3.2 Tehostetaan resursointia ja rahoituksen saatavuutta kansainvälisille markkinoille pääsyyn

- Kansainvälisille markkinoille pääsyyn liittyy H2020-työohjelma. Työohjelman sisältöön tuetusti vaikuttavien tahojen tulisi edistää Suomen kokonaiskilpailukykyä.
- Tutkimuslaitokset on syytä mainita korkeakoulujen rinnalla. Yliopistot sisältyvät jo korkeakoulu -käsitteeseen, joka kattaa sekä yliopistot että ammattikorkeakoulut.

”Vaikutetaan systemaattisesti H2020 tulevaan (2018-2020) työohjelmaan, jotta se tukee eurooppalaisen huipputaiteen kehittymistä alalla ja suomalaisen osaamisen hyödyntämistä ja kilpailukykyä (TEM, TEKES, VTT, CSC, Suomen Akatemia, korkeakoulut ja yliopistot, kehitysyhtiöt).”

- CSC:n on mainittu listassa ainoana yrityksenä. Tulisiko tässä puhua koko yrityskehittäjästä eikä vain yhdestä sen edustajasta? Tekstiehdotus:
”Vaikutetaan systemaattisesti H2020:n tulevaan (2018-2020) työohjelmaan, jotta se tukee eurooppalaisen huipputaiteen kehittymistä alalla ja suomalaisen osaamisen hyödyntämistä ja kilpailukykyä (TEM, OKM, Suomen Akatemia, TEKES, VTT, korkeakoulut ja yliopistot, yrityskehittäjä).”

Linjaus 3.3 Vaikutetaan kansainvälisellä yhteistyöllä datatalouden kehityssuuntiin

- Tähän on hyvä saada konkretiaa. Liian ympäröivää ilmaus ei tue yritysten kehittymistä. Toimijoiden yhteistyötä on syytä edistää mm. EU-tasolla esim. Public-Private Partnership -lähtökohdasta.