



Liikenne- ja
viestintäministeriö

Big data- strategia

Sisällysluettelo

1.	Johdanto	2
1.1	Big datan aikakausi	2
1.2	Miksi big data on niin tärkeää?	3
1.3	Kansainvälinen kehitys	5
1.4	Suomen positio big data-kilpailussa	5
1.5	Strategian lähestymistapa	7
2.	Edellytykset	7
2.1	Datatietoisuus	8
2.2	Koulutus	9
2.3	Soveltava osaaminen	11
2.4	Tutkimus	12
2.5	Infrastrukturi	14
2.6	Tekniset käytännöt ja standardit	15
2.7	Datan käytettävyys ja avoimuus	16
2.8	Lainsäädäntö, sääntely ja tietosuojat	18
2.9	Yhteistyö ja datan vaihto	19
2.10	Kokeilut ja t&k rahoituksen suuntaaminen	21
2.11	Oman tiedon hallinta	21
3.	Sovellusalueet ja niiden potentiaali Suomessa	22
3.1	Terveys	22
3.2	Älykkäät verkot	23
3.3	Tutkimus	25
3.4	Liikenne	25
3.5	Teollinen internet	25
3.6	Cleantech	27
3.7	Markkinointi ja mainonta	28
4.	Julkishallinto big datan sovellusalueena	28
4.1	Suorituskykyä parhailla käytännöillä	30
4.2	Kansalaisille yksilöllisempää julkista palvelua	31
5.	Läpileikkaavia teemoja	34
5.1	Mydata	34
5.2	Quantified self	34
5.3	Joukkoistaminen	35
5.4	Tiedon etsintä	36
5.5	Paikkatieto	36
6.	Toimenpiteet	37
6.1	Datatietoisuus	37
6.2	Koulutus	38
6.3	Soveltava osaaminen	39
6.4	Tutkimus	39
6.5	Infrastrukturi	39
6.6	Tekniset käytännöt ja standardit	40
6.7	Datan käytettävyys ja avoimuus	40
6.8	Sääntely	41
6.9	Yhteistyö ja datan vaihto	41
6.10	Kokeilut ja t&k rahoituksen suuntaaminen	43
6.11	Mydata	44
7.	Vaikuttavuus	44
8.	Strategian toimeenpano	46

1. Johdanto

Liikenne- ja viestintäministeriö asetti vuoden alussa big datan käyttö-työryhmän. Työryhmä on laajapohjainen; mukana on edustajia niin hallinnosta kuin suurista ja pienistä yrityksistä.

Työryhmä on kokoontunut kevään ajan ja valmisteellut kansallista big data-strategiaa. Työryhmän työn pohjalta strategialuonnoksessa esitetään kuvaa big datasta Suomessa sekä potentiaalisista, oleellisista ja Suomen kannalta tärkeistä sovellusalueista. Työryhmä on pohtinut myös tarvittavia edistämistoimia, jotta big data-ratkaisujen käyttö ja niistä saatavat hyödyt yleistyisivät Suomessa.

Työryhmän työ ja kansallinen big data-strategia valmistuu 30.6 mennessä.

1.1 Big datan aikakausi

4 00 000 000 000 000 000 000 tavua eli 4,4 zettatavua oli tutkimusyhtiö IDC:n mukaan eri laitteiden luomasta ja tallennetusta digitaalisen datasta muodostuvan niin sanotun digitaalisen avaruuden koko vuonna 2013.ⁱ Usein puhutaankin siitä, että olemme siirtyneet massiivisten tietoaaineistojen, niin sanotun big datan, aikaan.

Mistä big dataa oikein syntyy? Pääasiassa verkkoon kytketyistä laitteista, niiden lähettämästä ja erilaista sensorijärjestelmistä, sosiaalisesta mediasta, erilaisista verkon yli tehtävistä transaktioista, perinteisestä mediasta, tietojärjestelmistä, yritysten liiketoimintaan liittyvistä ohjaus- ja raportointijärjestelmistä, erilaisista digitaalisista arkistoista ja rekistereistäⁱⁱ Dataa syntyy kiihtyvällä vauhdilla lähes kaikesta internetiin jollakin tavalla liitetystä asiasta ja järjestelmästä.

Lopputuloksena on se, että erilaisten laitteiden, palveluiden ja sensorijärjestelmien yleistymisen ja lisääntyvän käytön seurauksena digitaalinen avaruus kasvaa vuosittain 40 %:lla eli kymmenkertaiseksi vuoteen 2021 mennessä.

Syntyvä data on muodoltaan hyvinkin erilaista, aina yksinkertaisista tekstidokumenteista ja sosiaalisen median päivityksistä reaaliaikaisen videovirran ja verkkoon kytketyn hissien sensoritiedon kautta hyvinkin monimutkaisten satelliittijärjestelmien tuottamaan dataan. Ennen valtavan ja erimuotoisen tietomäärän osittainen tallentaminen, yhdistely ja analysointi ei ollut mahdollista.

Teknologisen kehityksen ja siihen liittyen etenkin pilvipalveluiden kehityksen seurauksena digitaalisen tiedon tallennus- ja prosessointikustannukset ovat laskeneet jopa sadasosiin kymmenen vuoden takaisista kustannuksista. Nämä kehitystrendit yhdessä uusien digitaalista tietoa käsittelevien laskentamenetelmien kehityksen kanssa ovat tuoneet suurten data-aineistojen hyödyntämismahdollisuudet myös muiden kuin suuryritysten ulottuville. Hyödyntämismahdollisuuksien etsintää ja kehitystä on lisännyt vielä tiedon viestintäteknologioiden ja viestintäverkkojen muassa tuoma reaaliaikaisen käsittelyn ja siihen liittyvän päätöksenteon – yhä useammin automaattisen päätöksenteon - reaaliaikaisuuden mahdollisuus.

Yhteisesti jaettua ja täysin yksiselittäistä määrittelyä ei termille "big data" ole. Useimmiten käytetään kolmen V:n määritelmää, jolla viitataan sekä datan määrään (volume), syntyvauhtiin (velocity) sekä erilaiseen, struktoroimattomaan dataan (variety). Eri näkökulmista, muun muassa eri palveluntarjoajien määritelmässä, painotetaan ilmiön eri puolia. Esimerkiksi Intel määrittelee big datan puhtaasti koon (mediaaniarvona 300tb dataa viikottain), Microsoft lähestyy asiaa analyysin vaadittavien monimutkaisten menetelmien kautta. Lisäksi termiä voitaisiin lähestyä myös analyysiin käytettävien työkalujen kautta. (Mm. Hadoop)

Big Datan voi käsittää myös tietynlaisena tiedon paradigmanmuutoksena. Siirrytään yrityksissä ja hallinnossa tekemään päätöksiä suoraan "mitattuun" tietoon perustuen. Tutkimuksessa ei välttämättä tarvitse samalla tavalla muodostaa teoriaa, kun voidaan käydä valtavia tietomassoja läpi ilman ennakko-oletuksia ja ns. "löytää" malleja, yhteyksiä ja selityksiä.

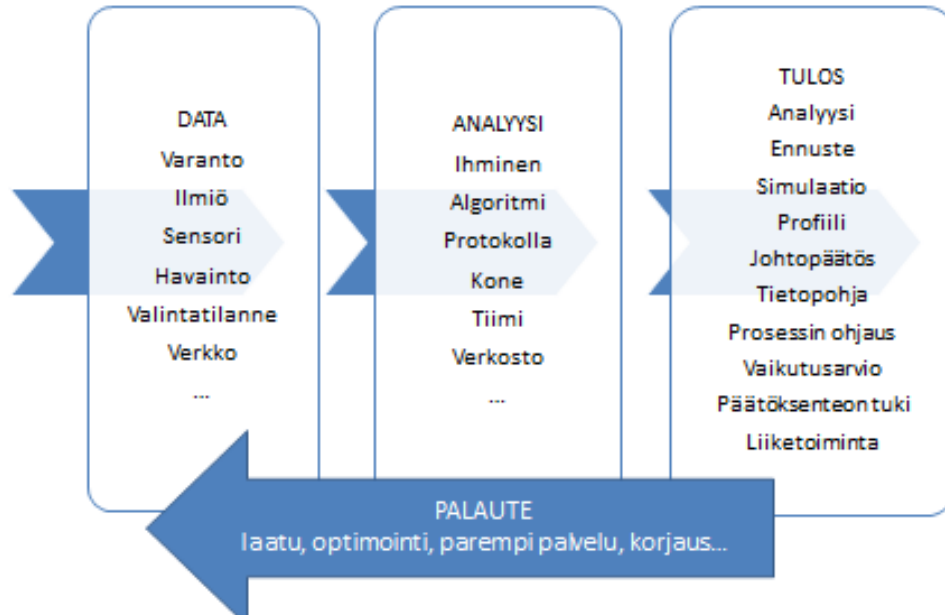
Erittäin suurten tietomassojen hyödyntämistä koskevan yhteiskunnallisen ja taloudellisen keskustelun yhtenä merkkipaaluna voidaan pitää konsulttiyhtiö McKinseyn vuonna 2011 julkaisemaa tutkimusta "Big data: The next frontier for innovation, competition and productivity".ⁱⁱⁱ Tutkimus arvioi big datan hyödyntämisen potentiaalisen vuosittaisen arvon Yhdysvaltojen terveydenhuollolle 300 miljardiksi dollariksi ja EU-alueen julkiselle sektorille 250 miljardiksi euroksi. ' Poliittisesti huomionarvoista oli myös tutkimuksen lupaus uusista työpaikasta: Pelkästään Yhdysvalloissa töitä on luvassa 140 000 – 190 000 data-analyytikolle. McKinseyn ohella myös muut suuret tutkimus- ja konsulttiyhtiöt pitävät big dataa yhtenä lähivuosien tärkeimmistä teknologisista trendeistä.

1.2 Miksi big data on niin tärkeää?

Big datan laajemman hyödyntämisen avulla haetaan useita selkeitä etuja, niin yritysmaailmassa kuin julkisen hallinnon puolella. Näihin mahdollisuuksiin ja hyötyihin kuuluvat muun muassa toiminnan optimointi ja siitä seuraavat säästöt, tarkemman tiedon saaminen päätöksenteon tueksi, selkeästi tarkemman tilannekuvan saaminen, uusien mallien ja yhteyksien löytäminen, parempi asiakaspalvelu sekä parhaimmillaan myös tulevan ennustaminen.

Oheiseen kuvaan on tiivistetty big datan lähteitä ja keruuta, data-analyysin osia sekä tavoiteltuja tuloksia eli hyötyjä. Saadut tulokset ja hyödyt vaikuttavat taas takaisin datan kerääntymiseen ja keräämiseen.

Datan hyödyntäminen



Toisaalta big datan aikakausi tuo myös vakavia uhkia. Näihin kuuluvat muun muassa yksityisyyden ja luottamuksen vaarantuminen, (usein henkilöitä koskevan) tiedon omistajuuden keskittyminen vain harvoille toimijoille, liiallinen käyttäjien seuraaminen ja liiallisesta profiloinnista mahdollisesta seuraava diskriminointi, esimerkiksi erilaisten maksujen suhteeton kallistuminen profiiliin perustuen. Tämä taas voi lisätä eriarvoisuutta ja eriytymistä. Tämä on ongelmallista varsinkin jos tähän johtaneet päätelmät on tehty virheellisen tiedon tai mallien perusteella. Big datan käyttö vastuuttomasti ja ymmärtämättä voi johtaa myös täysin virheellisiin kategoriointeihin. Tulkinnessa on myös muistettava, että data voi olla painottuneesti valikoitunutta, varsinkin kun kyse on henkilöistä kerättävästä datasta.

Tietomäärät ja big datan uudet mahdollisuudet eivät sinänsä poista tiedon tulkinnan merkitystä. Tiedon hyödyntäminen vaatii ymmärrystä esimerkiksi tilastollisesti päättelystä.

Digitalisaation ja big datan aikakaudella, lähes kaiken kytkeytyessä verkkoon, myös tieto- ja kyberturvallisuus sekä luottamus nousee ylipäänsä aivan keskeiseksi asiaksi. Organisaatioiden ja samalla koko yhteiskunnan toimivuus on pitkälti kiinni luotettavasta tietoturvasta.

Samalla big data-kehitys tuo myös aivan uusia haasteita tietoturvan ja luottamuksen kannalta. Valtavia tietomassoja ei voida aina siirtää, vaan analyysi tehdään suoraan tiedon tallennuspaikassa. Ulkopuolinen analyysisovelluksen tulee siis päästä suoraan järjestelmään. Niinpä eteen tulee tilanteita, joissa tietoa hallussapitävä organisaatio joutuu päättämään, luottaako se ulkopuoliseen sovellukseen niin paljon, että päästää sen

suoraan sisään järjestelmäänsä ja käsiksi tietoaaineistoihinsa. Erilaisia käytäntöjä täytyy miettiä tietoturvan näkökulmasta äärimmäisen tarkasti.

1.3 Kansainvälinen kehitys

Euroopan ja Suomen kehityksen katsotaan tulevan muutaman vuoden jäljessä kansainvälisesti suuria big datan hyödyntäjämaita, erityisesti USA:ta.

Strategioiden tarkastelussa huomaa, että eri maissa alan kehitykseen sekä muun muassa säästöjen ja kilpailukyvyn kannalta big data nähdään yhtenä oleellisimmista kehityskuluista. Niinpä monissa maissa alalle laitettut panostukset ovat mittavia. Big data-kehityksen kärjessä kulkeminen nähty äärimmäisen tärkeänä. Big data-maailmassa kilpailu on todella kovaa ja investoinnit alalle suuria.

USA:ssa Obaman hallinto julkaisi toukokuun alussa uuden big dataa käsittelevän raportin. Raportissa big dataa tarkastellaan niin hallinnon kuin yritystenkin kannalta. Erityinen huomio raportissa on big dataan liittyvissä politiikkakysymyksissä. Erityishuomio kohdistetaan yksityisyyden suojaan. Tässä yhteydessä on tärkeä huomata, että eri maiden erilaiset lainsäädäntökehikot, jotka voivat osaltaan aiheuttaa haasteita big datan hyödyntämiseen liittyen.^{iv} Alan kärkimaassa Yhdysvalloissa on panostettu pelkästään hallinnossa big data-kehitykselle satoja miljoonia dollareita. Big data-alalla toimivien yritysten suhteen maalla taas on selkeä etumatka esimerkiksi Eurooppaan verrattuna.

Englannissa istuva hallitus on asiantuntija-arvioiden pohjalta määritellyt big datan yhdeksi kahdeksasta maan taloudellisen tulevaisuuden kannalta keskeisistä teknologiasta. Tähän liittyen hallitus julkaisi vuonna 2013 myös huomattavia investointilupauksia sisältävän toimenpideohjelman ”Seizing the data opportunity – A strategy for UK data capability”.^v

Datan hyödyntämismahdollisuudet ja big data ovat nousseet vahvasti esille EU:n Digitaaliseen Agendaan ja uuteen Horizon –ohjelmaan liittyvissä hankkeissa. Samoin on laita OECD:n toiminnassa ja erityisesti New Sources of Growth: Knowlledge-Based Capital – teemaan liittyvissä projekteissa ja selvityksissä.

Saksassa ja Ranskassa big data nähdään myös tiiviisti teollisuuden kilpailukyvyn avaintekijänä. Maat ovat huomioneet laajasti big datan teollisuuspolitiikassaan. Ranskassa big data kuuluu seitsemän tärkeimmän tulevaisuuden teknologian joukkoon. Strategioissa nähdään big datan tuovan hallinnolle pidemmällä tähtäimellä merkittäviä kehitys- ja säästömahdollisuuksia.^{vi}

1.4 Suomen positio big data-kilpailussa

Digitaalisaation ja sen merkittävänä osana big datan mahdollisuuksiin on luonnollisesti avainkysymys myös suomalaisen elinkeinoelämän ja koko yhteiskunnan tulevalle kehitykselle. Big dataan liittyvät disruptioilmiöt yhtäältä ravistelevat myös suomalaisia

yrityksiä ja niiden toimintaympäristöä, mutta toisaalta avaavat myös uusia mahdollisuuksia kilpailukyvyyn, kilpailemaan ja uusien liitetoimintojen kehittämiseen.

Suomella hyvät edellytykset vastata big datan haasteeseen: Kansainvälisisten vertailututkimusten mukaan Suomi on maailman digitaalisesti edistyneimpien maiden joukossa erityisesti mitä tulee teknisiin valmiuksiin. Suomessa on esimerkiksi korkeatasoiset digitaaliset perusrekisterit ja myös runsaasti osaamista, jonka päälle voidaan suhteellisen nopeasti rakentaa kilpailukykyistä big data -osaamista.

Yhtenä tämän strategian tehtävänä on nostaa esille sovellusalueita, joissa Suomella on selvästi potentiaalia löydettävissä datan hyödyntämisessä. Perinteisesti meillä on osaamista mm. lääketieteellisessä tutkimuksessa, mobiiliteknologioissa, peliteollisuudessa ja ympäristömonitoroinnissa, jotka kaikki ovat hyvin dataintensiivisiä ja sen monimuotoiseen analyysiin perustuvia aloja. Toisaalta strategian avulla on kohdistettava toimia myös niihin yhteiskunnallisesti ja taloudellisesti merkittäviin alueisiin, missä olemme selkeästi kehityksessä jäljessä, mutta joista olisi saatavissa selkeitä hyötyjä.

Suomella on myös vahvaa menetelmä- ja it-osaamista, jota muuntamalla ja hyödyntämällä koulutuksen, tutkimuksen ja asiantuntemuksen jakamisen kautta saataisiin mukaan big data kehitystyöhön.

Big datan hyödyntäminen julkisella sektorilla on vasta alkutekijöissään, mutta tarjoaa suuria mahdollisuuksia niin palvelujen kuin prosessienkin parantamiseen ja tehostamiseen sekä uusiin toimintatapoihin. Suomi on ollut edelläkävijämaita avoimessa datassa ja julkinen sektori avaa tietoaaineistojaan vaihtelevaan tahtiin. Tätä avoimuuden ja hyvälaatuisten julkisten tietovarantojen saatavuuden kulttuuria tulisi hyödyntää myös big data kehitystyössä. Avoimen ja big datan, julkisten ja ykistyisten data-aineistojen yhdistelyssä piilee todella suuri potentiaali.

Big datassa on huomioitava erityisesti erilaisten datojen vaatimukset datan käytön ja hallinnan suhteen. Esimerkiksi henkilötietona terveysdatan käsittely tarjoaa erilaisia käyttömahdollisuuksia kuin esimerkiksi laitteista saatava sensoridata (IoT). Suomi on hyvin korkean tietosuojaan maa. Nykyisessä maailmassa tätä voidaan pitää uudenlaisena vahvuutena, jota ei saa vaarantaa. Toisaalta korkealaatuinen viestintäinfrastruktuuri ja Suomen vahvuus mobiilipuolella antaa hyvän peruspohjan erilaiselle teolliseen internetiin ja internet-of-thingsiin liittyvälle kehitykselle. Tieteellisessä laskennassa ja tiedon käsittelymenetelmien osaamisessa on myös yksi Suomen vahvuus.

Toisaalta hitaalla reagoinnilla saatetaan vahvuudet hukata. Niin hallinnossa ja yrityksissä täytyy tehdä tarvittavia muutoksiam, jotta uusia ratkaisuja voidaan kehittää ja ottaa käyttöön. Muualla big dataan panostetaan jo vahvasti. Vaarana on kehityksen keltasta putoaminen.

Suomessa aiempiin big dataan liittyviin kansallisiin toimenpiteisiin kuuluu muun muassa ICT 2015-ryhmän polku 10.^{vii}

1.5 Strategian lähestymistapa

Tämän strategian lähestymistapa big dataan lähtee erilaisista ratkaisuksista ja kehityksestä saatavan hyödyn kautta. Strategian lähestymistavassa korostetaan erityisesti:

- hyödynnettävyyttä,
- usein suurimman hyödyn tuovaa monialaisuutta, poikkihallinnollisuutta ja -tieteellisyyttä,
- datan avoimuutta sekä avoimen datan ja big datan yhdistelyä.

Strategiassa big dataa ei ole lähdetty määrittelemään liian tiukasti esimerkiksi pelkän koon, teknologisten ratkaisujen tai käsittelymenetelmien kautta.

Muun muassa teknologioiden, analyysimenetelmien ja sensoritekniikan kehittyessä datan keruu, siitä tehtävät havainnot ja yleisesti koko big data-kehitys on vielä murrosvaiheessa. Tulevaisuutta ja tulevia uusia ratkaisuja on vaikea ennustaa. Niinpä tässä strategiassa keskitytään pitkälti Suomen nykytilan kuvaukseen ja yleisiin edellytyksiin, joilla kilpailussa pysytään mukana.

Uusien big data-menetelmien tuomat mahdollisuudet ja hyödyt vaativat usein monialaisuutta ja poikkihallinnollista sekä -tieteellistä lähestymistapaa. Lähtökohtana nähdään tarve laajalle yhteistyölle. Data-aineistot sijaitsevat erilaisten organisaatioiden ja instituutioiden huomassa, joiden yhdistelystä ja soveltamisesta syntyy arvoa. Laajan yhteistyön tarve kohdistuu siten niin hallinnon, yritysten kuin esimerkiksi järjestöjen ja kansalaisten kesken.

Strategiassa on kiinnitetty tähän erityishuomiota. Tiedon hyödyntäminen big data-menetelmien avulla vaatii useimmiten monen erilaisen ja perinteisesti eri alalle kuuluvan tiedon käyttöä ja yhdistelemistä. Samalla big datasta saadaan suurinta hyötyä, mikäli erilaista tietoa on käytettävissä. Yksi strategian painotuksista koskeekin avointa dataa. Tietoa tulisi olla saatavilla ja käytettävissä mahdollisimman avoimesti, toki esimerkiksi yksityisyyden suojaa vaarantamatta.

Strategiassa ei siis keskitytä pelkästään massiivisten, esimerkiksi supertietokoneita analyysiin vaativiin reaaliaikaisiin tietomassoihin. Strategia kohdistuu myös käsittelymenetelmällisesti yksinkertaisempaan dataan ja erityisesti erilaisten tietoa-aineistojen yhdistelyyn. Ajatuksena on ollut tunnistaa ja löytää Suomen kannalta oleellisia alueita, joilla kerättävä tieto, uudet tiedonkäsittelymenetelmät ja laajempi tiedon hyödyntäminen voivat tuoda arvoa. Strategian tavoitteet ja toimet kohdistuvat niin yksityiseen kuin julkiseen sektoriin.

2. Edellytykset

Tässä luvussa tunnistetaan ja kuvataan big datan käytön ja sen yleistymisen kannalta oleellisia yleisiä edellytyksiä. Edellytysten tulee olla kunnossa, jotta big data-ratkaisuja voidaan soveltaa eri aloille ja saavuttaa hyötyjä.

2.1 Datatietoisuus

Datatietoisuudessa on kyse pitkälti datan potentiaalisen arvon ja toisaalta riskien ymmärtämisestä. Hyöty ei aina näy heti ja ole saatavissa suoraviivaisesti, mutta tulevaisuudessa ja esimerkiksi muiden, esimerkiksi reaaliaikaisten datalähteiden kanssa yhdistelemällä saatetaan saavuttaa huomattavia etuja. Kyse on siis datan itsensä arvon ymmärtämisestä. Erilaisesta datasta saadaan oleellista informaatiota ulos, usein kyse on siitä, että osataan vain kysyä oikeita kysymyksiä ja hakea oikeita asioita. Datatietoisuus koskee niin yrityksiä, hallintoa kuin yksilöitäkin.

Nykyaikaisen analytiikan ja big datan omaksuminen osaksi jokapäiväistä työtä on keskeinen askel kohti tiedolla johtamisen yrityskulttuuria. Uuden teknologian ja prosessointikapasiteetin kustannusrakenteiden jatkuva kehitys vaatii omien kyvykkyyksien säännöllistä uudelleentarkastelua. Se mikä datan tarkastelussa vuosi sitten oli mahdotonta tai kannattamatonta, voi tänään olla varsin perusteltua ja liiketoiminnallisesti kannattavaksi todistettavaa.

Samalla budjetilla voidaan ensi vuonna jälleen tarkastella laajempaa datamassaa, etsien tarkempia signaaleja, uusia toimintamahdollisuuksia ja aiemmin tunnistamattomia optimointikohteita. Tämä sykli toistuu yhä uudelleen ja organisaatioiden onkin opittava kyseenalaistamaan vanhat datan tarkastelun rajat ja kiinnittämään tarkemmin huomiota uusien lähteiden syntyyn. On tultava datatietoiseksi. Maailmalla on tietopohjaisesta päätöksenteosta esimerkkinä mm. amerikkalainen autonvalmistaja Ford, joka selvisi autoteollisuuden kriisistä siirtymällä data-driven päätöksentekoon.^{viii}

Perinteisesti organisaatiot ovat tottuneet tarkastelemaan oman liiketoimintansa tuottamaa dataa. Ja siitäkin erityisesti esivalittua osajoukkoa, liiketoiminnan ydintietoa. Tämä tieto on kerättyä erilaisissa perusjärjestelmissä (ERP, CRM, CMS) ja sekä datan tuntemus että sen analysointiin käytettyjen välineiden osaaminen on vahvaa. Big datan huomiointi tässä kontekstissa tarkoittaa datan keräämisen laajuuden tarkastelua.

Jos perusjärjestelmien lokitasot, tapahtumien kirjaaminen tai tapahtumien sisällön laajuus on aiemmin rajoitettu tallennus- ja prosessointikapasiteetin kustannusten perusteella, voidaan näitä arvioida uudelleen nykyisen kustannusrakenteen ja kapasiteetin myötä. Ydinliiketoimintaa on myös syytä tarkastella säännöllisesti uudelleen tiedonkeräämisen potentiaalin näkökulmasta. Onko aiempi analytiikka tai toiminnan tarkastelu nostanut esiin asioita, joissa datan kerääminen ei ole riittävällä tasolla? Voitaisiko tuota tunnistettua, kiinnostavaa dataa kerätä tuomalla järjestelmiin uusia ominaisuuksia tai toimintaympäristöön esimerkiksi erilaisia sensoreita? Näihin mahdollisuuksiin tarttuminen kartuttaa oman liiketoiminnan tuottamaa datavarantoa entisestään ja saattaa nostaa big datan teknologioiden omaksumisen varsin ajankohtaiseksi.

Datatietoinen organisaatio ei kuitenkaan rajoita tiedonjanoaan ainoastaan omien järjestelmien keräämään dataan, vaan katsoo ympäröivää ekosysteemiä laajemmin. Datan mahdollisuuksia on syytä arvioida kumppaniverkoston kanssa, tunnistaa synergioita, datan vaihdon mahdollisuuksia tai jopa uuden liiketoiminnan perustamisen potentiaalia datavarantojen yhdistämisen myötä.

Hallinnossa big datan mahdollisuuksien ja mahdollisten riskien tunnistaminen on oleellista. Big dataan ja tiedon luovaan yhdistelemiseen perustuvien ratkaisuiden avulla voidaan saavuttaa muun muassa merkittäviä säästöjä, tehdä informoituneempia päätöksiä selkeästi parantaa kansalaisten palvelemista.

Lähes kaikkien hallinnonalojen tavoitteissa ja erilaisissa strategioissa on monia asioita, joissa luova datan ja uusien big dataan perustuvien ratkaisujen käyttö voi selkeästi parantaa toimintaa. Samalla eri puolilla hallintoa tulisikin tunnistaa, kuinka datalähtöisten ratkaisujen avulla voidaan etsiä helpotusta jopa suuriin yhteiskunnallisiin ongelmiin, kuten ikääntymiseen sekä harmaaseen talouteen.

Yksilöiden kohdalla datatietoisuudella tarkoitetaan muun muassa sitä, että ymmärrys omien tietojen jakamisen riskeistä on kohdallaan. Vaikka yksittäisellä tiedolla ei vielä välttämättä tee vielä mitään, uusia menetelmiä käyttämällä ja erilaisia tietoaaineistoja yhdistelemällä voidaan saada tarkasti selville, mitä kuka on kyseessä, miten hän on toiminut ja jopa miten hän tulee toimimaan. Tällaiset tiedot ovat kullanarvoisia monessa toiminnassa. Yksilön ymmärrys muun muassa tiedon ja sijainni jakamista, kerätystä tiedosta ja muun muassa profiloinnin erilaisista mahdollisuuksista ja riskeistä olisi hyvä olla olemassa. Samalla omien tietojen kontrolloidullakin ja tietoisella luovuttamisella yksilö voi saada monenlaisia hyötyjä, muun muassa parempaa palvelua.

Ylipäänsä tulee huomioida myös big datan käyttöön liittyvät mahdolliset riskit, esimerkiksi yksityisyydensuojaan liittyen. Kehityksessä ja erilaisissa toimenpiteissä täytyy pohtia myös mahdollisesti seuraavia ongelmia.

Esimerkiksi Sote-tieto hyötykäyttöön- strategiassa big data on tunnistettu tärkeäksi työkaluksi tavoitteiden saavuttamiselle. Myös usealla muulla hallinnonalalla ja varsinkin yhteisesti eri alojen välillä big data-ratkaisujen käyttöönotolla voitaisiin asetettuja tavoitteita saavuttaa huomattavasti tehokkaammin.

2.2 Koulutus

Big datan hyödyntämisen kannalta koulutus on avainasemassa. Tällä hetkellä muun muassa erilaisten analyysimenetelmien osaajista on pulaa. Pulaa on myös big datan käsittelyä osaavista ohjelmoijista.

Samoin tarvitaan yritysjohtajia, jotka ymmärtävät data-analyysin mahdollisuudet päätöksenteon tukena ja yritysten kilpailukyvyyn turvaajina. Johtajien tueksi tarvitaan henkilöitä, jotka osaavat muuntaa analyysien tulokset raporteiksi johtajille.

Tulevaisuudessa osaajilta vaaditaan yhä enemmän, jolloin koulutuspuutteet voivat muodostua merkittäväksi esteeksi big datan laajemmassa hyödyntämisessä.

Koulutuspuutteisiin täytyy alkaa vastaamaan jokaisella koulutustasolla, jo peruskoulusta lähtien sekä myös yritysten taholla esimerkiksi täydennyskoulutuksen muodossa. Perustana olevia matemaattisia, tilastotieteellisiä sekä ohjelmointitaitoja tulisi painottaa opetussuunnitelmissa jo hyvin varhain, jotta tulevat osaamistarpeet tulisivat paremmin katettua pitkällä tähtäimellä. Tiettyjen tärkeiden tietotyökalujen opetusta tulee esimerkiksi keskiasteella lisätä. Lisäksi datalähtöisempi ja empiirisen kokeilevaa lähestymistapaa tulee korostaa.

Tämä ei kuitenkaan vastaa välittömiin tarpeisiin. Suomessa on käynnistymässä sekä ammattikorkeakouluissa että yliopistoissa useita big dataan ja data-analytiikkaan liittyviä koulutusohjelmia. Yliopistotasolla esimerkiksi Aalto-yliopiston kaikkien kuuden koulun maisteriopiskelijoiden on syksystä 2014 lähtien mahdollista suorittaa sivuaineen laajuinen "Analytics and Data Science" -kokonaisuus.

Big Data-koulutusohjelmia Suomessa:

Jyväskylän yliopisto: Data-analyysin monitieteellinen maisterikoulutus (DATA)
<https://www.jyu.fi/it/laitokset/mit/opiskelu/maisteriopinnot>

Tampereen teknillinen yliopisto: Big Data ja data-analytiikka liiketoiminnan kehittämisessä- täydennyskoulutus.
<http://www.tut.fi/fi/yrityksille/taydennyskoulutus/koulutustarjonta/big-data/index.htm>

Aalto yliopisto: Master's Degree Programme in Information and Service Management
http://www.aalto.fi/fi/studies/education/programme/information_service_management_master/

Itä-Suomen yliopisto: Datatieteen koulutusohjelma
<https://www.uef.fi/fi/cs/datatiede>

Tampereen yliopisto: Masters Programme in Computational Big Data Analysis

Kaikkien suomalaisten korkeakoulujen, jotka eivät vielä ole havahtuneet big datan vaikutuksiin, tulisi huomioida kehitys ja pikaisesti saada datalähtöisyys ja data-analyysi näkymään eri alojen koulutuksessa. Näissä koulutusohjelmissa voitaisiin tehdä yritysten kanssa tiivistä yhteistyötä, jotta tulevat varsin akuutit tarpeet saadaan edes osittain katettua myös suoraan käytännön taidoilla varustetuilla valmistuneilla. Erilaisilla harjoittelu- ja tutustumismahdollisuuksilla on oma tärkeä osansa. Data-analytiikkaan koulutettaville tulisi tarjota mahdollisuuksia opiskelun aikana erilaisiin harjoitteluihin. Näitä voisi yritysten lisäksi tarjota myös julkishallinnossa, mikä osaltaan auttaisi siirtymistä datalähtöisempään toimintatapaan.

Koulutustarpeita Big data-alalla E-Skills UK:n selvityksen mukaan:

<http://www.e-skills.com/research/research-publications/big-data-analytics/>

Lisäksi alueella tulisi lisätä monialaista koulutusta. Datalta täytyy osata kysyä oikeita kysymyksiä, joten erilaisten taitojen ja erilaisen osaamisen yhdistäminen on välttämätöntä. Niinpä data-analytiikkaa ja datan "ymmärtämistä" tulisi näkyä melkein alan kuin alan koulutuksessa, aina kauppatieteellisestä yhteiskunta- ja sosiaalitieteisiin. Pelkkä tekninen osaaminen ei riitä, vaan tarvitaan ymmärrystä substanssialoista sekä kykyä yhdistellä asioita ja aloja luovasti. Erilaisia vahvuuksia sisällään pitävät monialaiset tiimit tulevat työskentelemään big datan parissa. Niinpä jo koulutuksessa erilaista

tiimityöskentelyä täytyy tukea ja erilaisen koulutustaustan omaavia saattaa yhteisten ongelmien eteen.

Monella alalla käytettävissä olevan tiedon hyödyntäminen toisi huomattavaa lisäarvoa. On tärkeää kouluttaa osaajia myös yksityisyyden ja tietosuojan alueilla, sillä EU:ssa valmisteilla oleva lainsäädäntö saattaa tuoda yrityksille merkittäviä sanktioita (mahdollisesti 5% liikevaihdosta) henkilötietojen virheellisestä käytöstä. Nykyinen lainsäädäntö antaa yrityksille mahdollisuuden käyttää dataa vain siihen tarkoitukseen kun se on alunperin kerätty, mikä luonnollisesti rajoittaa datan käyttöä huomattavasti.

Opettajien osaamisen rajoitteet voivat muodostaa big data-osaamisen kehitykselle haasteen, kaikilla koulutustasoilla. Yhtenä ratkaisuna voivat olla erilaiset verkko-oppimistyökalut sekä MOOCit (massive open online course). Näiden hyödyntäminen vaatii esimerkiksi yliopistoilta halua yhteistyöhön.

2.3 Soveltava osaaminen

Yrityksien ja muiden organisaatioiden big datan hyödyntämisen esteeksi saattaa muodostua big data-osaamisen puute. Niinpä yrityksiltä ja julkiselta sektorilta vaaditaan big datan kanssa nyt tai tulevaisuudessa työskentelevän henkilöstön täydennyskoulutusta. Ongelmana on jälleen osaavien kouluttajien pieni määrä. Myös yritysten osalta erilaisten MOOCien ja muiden verkkotyökalujen käyttö voi helpottaa akuuttia pulaa. Samalla big data-kehityksen hyödyntäminen vaatii jatkuvaa osaamisen päivittämistä.

Erilaisen big dataan liittyvän osaamisen tarve on suuri. Uusia työnkuvia muodostuu jatkuvasti. Kaikkea big datan hyödyntämiseen liittyvää ei voida kouluttaa samalle henkilölle. Erilaisia vahvuuksia sisällään pitävät monialaiset tiimit tulevat työskentelemään big datan parissa. Oheisessa kuvassa on esimerkki siitä, kuinka laajaa osaamista big datan hyödyntäminen vaatii ja kuvaus big data-tiimin erilaisista rooleista.

Esimerkki Big Data –tiimistä

Teknologiaosaajat	Dataosaajat	Substanssiosaajat
 Hadoop asiantuntija	 Tilastotieteilijä	 Alan 1. asiantuntija
 DBMS asiantuntija	 Algoritmi-osaaja 1.	 Alan 2. asiantuntija
 Ohjelmoinnin asiantuntija	 Algoritmi-osaaja 2.	 Tiedon/tulosten esittäjä
 Matemaatikko	 Visualisoija	 Tietotaiteilija

Edellisessä kappaleessa mainittu yhteistyö oppilaitosten kanssa ja muun muassa käytännön harjoittelut auttavat tulevaisuudessa osaajapulaan. Akuutti (seuraavat pari vuotta) osaajatarve täytyy kuitenkin ratkaista pääosin muilla keinoilla. Vaarana on pitkälti USA:n johtaman kehityksen kelkasta putoaminen, jos osaavaa työvoimaa ei ole erilaisiin tehtäviin saatavilla.

Koulutuspuolen ja sekä yritysten ja hallinnon yhteistyötä tarvitaan lisää. Esimerkiksi erilaisia harjoittelumahdollisuuksia sekä opinnäytepaikkoja pitää saada lisättyä. Nämä koskevat niin suoran käytännön, kuten koodaamisen sekä vaikkapa laajemmin data-analytiikkaan ja data managementiin keskittyviä opiskelijoita. Lisäksi aineistoja tulisi saada koulutuskäyttöön. Tällöin opiskelijat voisivat oppia empiirisemmin suoraan käytännön ongelmien parissa.

2.4 Tutkimus

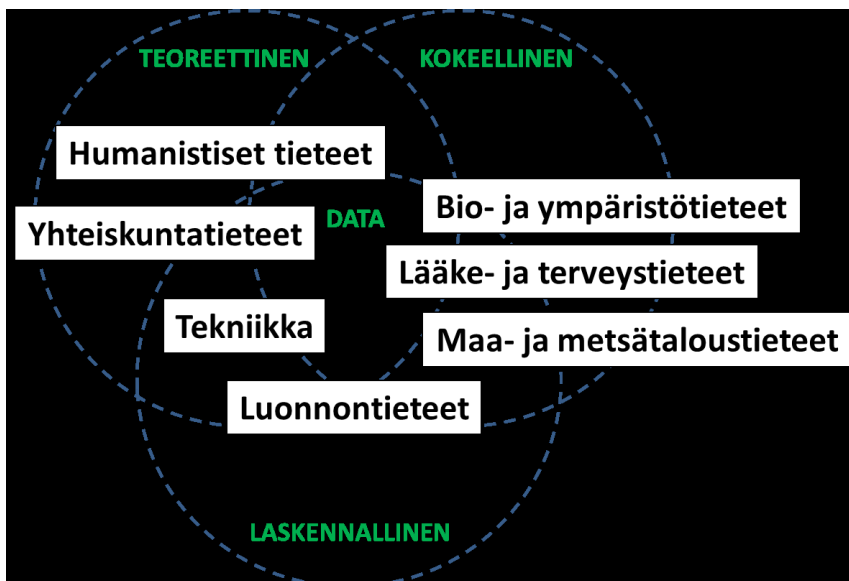
Big data tarjoaa kosolti mahdollisuuksia tutkimukselle eri tieteenaloilla. Samalla big data menetelmien ottaminen mukaan tutkimustoimintaan on osin välttämätöntä, tutkimustiedon ja esimerkiksi potentiaalisen tutkimusdatan määrän kasvaessa rajusti. Big data-menetelmien käyttö on suomalaisessa tutkimuksessa vielä jokseenkin vähäistä. Nykyiset big data-menetelmät ovat suurilta osin lähtöisin kaupalliselta puolelta. Tutkimuksessa tarvitaan myös erilaisia, nimenomaan tutkimuskäyttöön räätälöityjä menetelmiä.

Big dataa voidaan hyödyntää monilla tutkimusaloilla. Esimerkiksi biotieteiden, tähtitieteen, fysiikan ja vaikkapa psykologian (psykonometria) ja taloustieteen (ekonometria) alalla big data-kehityksestä ja menetelmistä on saatavilla selkeitä hyötyjä. Muun muassa näiden alojen tutkimukseen big data tarjoaa täysi uusia mahdollisuuksia.

Big dataa voidaan käyttää tutkimuksessa erilaisilla tavoilla. Samalla tutkijoiden ja tutkimusryhmien tarpeet ovat erilaisia. Parhaiden tulosten saaminen big datan avulla vaatii monialaista osaamista. Tätä voi kuvata muun muassa jakamalla big datan tutkimuskäyttäjät kolmeen ryhmään. Tietoteknisesti orientoituneet menetelmäkehittäjät keskittyvät enemmän tekniseen puoleen. Esimerkiksi tilastotieteilijätaustaiset data-analyytikot taas nimenomaan analyysimenetelmiin. Big datan tutkimukselliset soveltajat, esimerkiksi biologit, ovat taas lähinnä nimenomaan oman alansa asiantuntijoita. Yhteistyötä tarvitaan eri ryhmien kesken, jotta uusien big data-menetelmien hyöty saadaan tutkimuksessa ulosmitattua mahdollisimman laajasti.

Big dataan liittyy myös paljon yhteiskunnallisia kysymyksiä, joiden ymmärtämiseen tarvitaan monialaista tutkimusta. Digitaalisten aineistojen hyödyntäminen vaatii tuekseen viisasta sääntelyä. Tarvitaan yhteiskunnallista ja juridista ymmärrystä siitä, miten aineistoihin liittyvää lainsäädäntöä voidaan kehittää tavalla, joka vie eteenpäin toivottuja tapahtumainkuluja. Digitaalisesta jäljittämisestä pitää käydä avointa keskustelua, ja sen hyödyistä ja haitoista on oltava tutkimuksellisesti perusteltua näyttöä.

Tutkimusnäkökulmasta huomattavaa on se, että suuria aineistoja analysoitaessa yksilökohtaista aineistoa voidaan yhdistää lukuisista eri lähteistä ja toisaalta aineistot saattavat olla varsin epäyhtenäisiä. Ne voivat koostua eri paikkoihin jääneistä digitaalisista jäljistä tai olla kirjoitettuja viestejä, terveystietoja tai arkistoituja kuvia. Tällainen aineistojen moninaisuus haastaa käsitystä tiedon ja tutkimuksen luonteesta. Suurten aineistojen analyysia voi ajatella sarjana kokeiluja, jolloin onnistuneilla kokeiluilla voi kyseenalaistaa aiempaa osaamista ja kehitellä yhteistoiminnan muotoja, käytäntöjä ja sovelluksia. Näin kompetenssi rakentaa uutta tiedontuotannon maailmaa kasvaa merkittävästi.



Tutkimusnäkökulmasta big dataa voi ajatella myös yhteistoiminnan areenana. Tätä edellyttää monialaisia tutkimusryhmiä. Tietoa louhitaan yhteistyössä muiden kanssa ja aineistolle esitetään yhä uusia kysymyksiä. Kyse on jatkumosta, jossa kuljetaan pikkuhiljaa kohti mielenkiintoisia vastauksia tai aineiston esittämisen tapoja. Vahvat toimijat, kuten

Google, Amazon, tai NSA, hyödyntävät digitaalista aineistoa, jotta ne oppisivat tunnistamaan ihmisten aikeita, tarpeita ja liikkeitä. Yksittäisille ihmisille hyöty on huomattavasti rajallisempi kuin silloin, kun aineistoanalyysi aidosti tukee ihmisiä esimerkiksi heidän hyvinvointipyrkimyksissään tai arvoprojekteissaan. Tästä näkökulmasta olisi tärkeää, että aineistojen analyysimenetelmät kehittyisivät tavoilla, joilla ne myös hyödyntäisivät digitaalisia jälkiä jättäviä ihmisiä ja yhteiskuntaa laajemmassakin mielessä. Big data haastaa tutkijoita sekä käyttämään mielikuvitustaan että tekemään eettisesti kestäviä valintoja tutkimustyössään.

Tutkimusrahoitusta on ohjattava tavalla, joka suosii teknologiaosaajien ja yhteiskuntatieteilijöiden yhteistyötä big dataan liittyvissä asioissa. Eettisiä kysymyksiä ei voi ajatella teknologian käytöstä irrallisena alueena vaan ne on pidettävä mukana tutkimuksen kaikissa vaiheissa.

Suomessa big data-alan tutkimus on vielä suhteellisen hajanaista. Erilaisten, mielellään monialaisten big data-tutkimuskeskusten luonti voisi vauhdittaa alan kehitystä. Perinteisesti Suomessa on huippuluokan osaamista esimerkiksi tieteellisestä laskennasta (CSC), algoritmiikasta, tiedon louhinnasta, tilastollisista menetelmistä, tiedonhallinnasta ja näihin liittyvistä ohjelmointimenetelmistä. Big data -tutkimus vaatii näiden osaamisalueiden menestyksellistä ja luovaa yhdistämistä erilaisiin sovellusalueisiin. Suomalaisella big data-tutkimuksella on oivat mahdollisuudet kuulua maailman kärkikastiin

Useat maat pyrkivät big data-tutkimuksen ykköskaartiin. Maailmalla on perustettu muun muassa big dataan keskittyneitä tutkimuskeskuksia. Britit investoivat hiljattain Alan Turingin nimeä kantavan instituutin perustamiseen <https://www.gov.uk/government/news/plans-for-world-class-research-centre-in-the-uk>
<http://www.theguardian.com/technology/2014/mar/19/budget-2014-alan-turing-institute>

Suomessa huippututkimusta löytyy muun muassa Suomen Akatemian huippuyksiköistä. ALGODAN on keskittynyt algoritmiseen data-analyysiin, COIN data-analyysimenetelmiin, LASTU taas yhdistää menetelmien kehittämisen laskennallisesti vaativien ja tieteellisesti erittäin merkittävien kysymysten tutkimiseen.

<http://www.aka.fi/lastu#Lastu>
<http://www.cs.helsinki.fi/research/algodan/>
<http://research.ics.aalto.fi/coin/>

2.5 Infrastrukturi

Verkkoon kytkeytyneiden laitteiden ja niiden lähettämän sekä vastaanottaman datan määrä kasvaa huimaa vauhtia. Cison ennusteiden mukaan vuoteen 2017 mennessä yhteensä noin 19 miljardia laitetta on kytkettynä internetiin.^{ix} Suurin osa dataliikenteestä tulee tällöin muodostumaan koneiden keskinäisestä (machine-to-machine) viestinnästä.

Tähän liittyvät ip-ongelmat on pitkälti ratkaistu ipv6-protokollan käyttöönoton myötä. Osoitteiden riittävyys on käytännössä varmistettu uuden protokollan myötä.

Kaiken internetissä myös tulevaisuuden verkoilta, niin kiinteiltä kuin langattomilta, vaaditaan yhä enemmän^x. Pienilläkin viiveillä voi olla ratkaisevan suuri merkitys erilaisten big dataan perustuvien ratkaisujen kannalta. Big datan analyysi ja erilaiset ratkaisut perustuvat usein reaaliaikaisen tiedon käyttöön. Niinpä pienetkin katkot ovat kriittisiä. Big datan käytön kannalta verkon toimintavarmuus on äärimmäisen tärkeää.

Big datan hyödyntämisen kannalta olennaista on, että verkot toimivat hyvin yhteen niin kotimaassa kuin kv. datavirtojen kohdalla, ne ovat turvallisia, tehokkaita ja toimintavarmoja. Oleellisessa osassa ovat myös erilaiset yhteentoimivuutta varmistavat toimet.

Datan varastoinnin osalta datakeskusten tulee toimia mahdollisimman tehokkaasti. Suomen kannalta mahdollisuutena on varmistaa huippuluokan yhteyksien avulla uuden dataintensiivisen liiketoiminnan sijoittuminen maahan.

Langattomien verkkojen osalta tulee ennakoivasti allokoita riittävästi ennakoon mobiilidatakäyttöön soveltuvia taajuuksia. Verkkojen tulee kehittyä älykkäämmäksi, jotta valtaisten datamassojen siirtyminen on mahdollista. Verkkoon kytkeytyneiden laitteiden määrän kasvaessa mahdollisia ovat myös täysin laitteiden väliselle viestinnälle ja sensoritietolle tarkoitettut "internet of things"-verkot. San Franciscoon ja Piilaaksoon on jo tulossa tällaiseen tarkoitukseen allokoitu mobiiliverkko.^{xi}

2.6 Tekniset käytännöt ja standardit

Eri organisaatioihin kertyvä data täytyy pystyä ottamaan tehokkaasti käyttöön siellä missä sitä tarvitaan; tätä tarkoitusta varten tarvitaan yhteisiä sopimuksia tai standardeja siitä, kuinka dataa tallennetaan, jaellaan ja prosessoidaan. Big datan täysimääräisessä hyödyntämisessä oleellista on myös se, että erilaisista tietokannoista ja dataseteistä voidaan hakea tietoa, malleja ja tuloksia. Toisin sanoen yhteentoimivuus on oleellista.

Tällä hetkellä tätä edistäviä yhteisiä, maailmanlaajuisessa käytössä olevia, esimerkiksi datan varastointiin ja tiedonsiirtämiseen liittyviä big data-standardeja ei kaikilla osaluilla ole, teknologioiden yhä kehittyessä. Standardisointi kehitetään eri puolilla kovaa vauhtia, asiaa selvitetään muun muassa Yhdysvaltain National Institute of Standards and Technologyssa. Tämän lisäksi useissa muissa standardoimisorganisaatioissa pohditaan ja etsitään mahdollisuutta eri teknologioiden yhteentoimivuuden varmistaville big data-standardeille.

Samalla standardointia tarvitaan myös yritysten keskinäisen datan vaihdon mahdollistamiseen. Erilaiset big data-ratkaisut ja teknologiat eivät ole tällä hetkellä yhteenvoimivia. Tulevaisuudessa nähtäneen yhteisiä standardeja ja käytäntöjä, joiden avulla big data-ratkaisujen yhteentoimivuus varmistetaan.

Suomen ja suomalaisten yritysten tulee pysyä mukana ja edistää standardikehitystä osallistumalla mahdollisuuksiensa mukaan erilaisten ja eri alojen standardointielinten toimintaan. Lisäksi tietoa standardien ja teknologian kehityksestä tulee jakaa keskinäisessä verkostossa. Standardien ollessa vielä kehittymättömiä, niiden

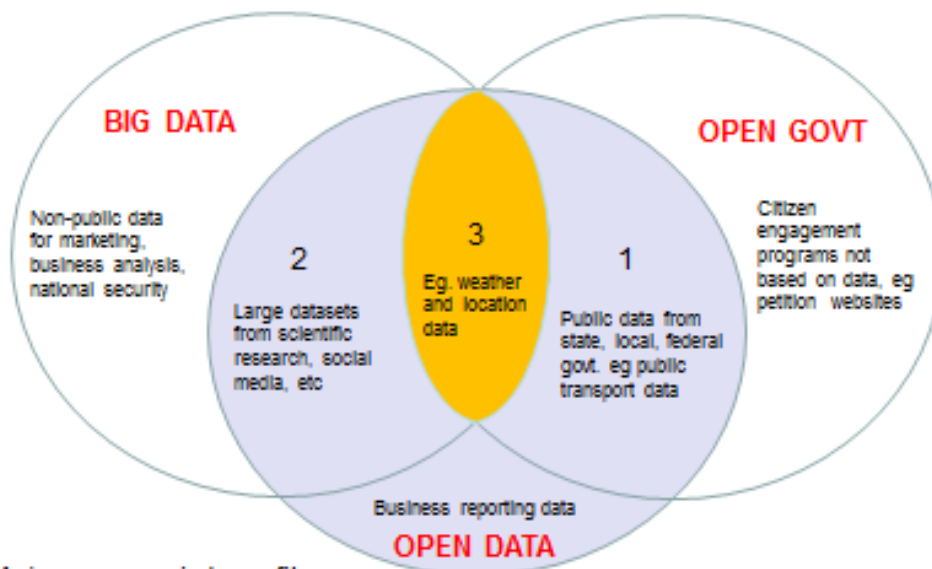
muotoutumiseen pyritään vaikuttamaan niin, että muotoutuvat standardit ovat mahdollisimman pitkälle suomalaisten alan yritysten edun mukaisia.

Suomessa on tehtävä myös big data-alan teknologiakartoituksia ja tunnistettava oleellisia teknologioita, malleja sekä standardeja. Samalla erilaisten big dataan liittyvien määritelmien ja termien yhtenäisyys on oleellista. Tätä tulisi myös Suomessa tehdä. (esim Standardoimisliitto)

2.7 Datan käytettävyys ja avoimuus

Yksi tapa lähestyä datan käytettävyyttä ja avoimuutta on esitetty oheisessa kuvassa^{xii}. Big dataa on olemassa ilman avointa dataa ja toisinpäin. Kuitenkin hyödyntämisen kannalta mielenkiintoisimpia alueita ovat big datan ja avoimen datan risteyskohdat.

Data with the most economic benefits



3 Major economic benefits

<http://www.theguardian.com/public-leaders-network/2014/apr/15/big-data-open-data-transform-government>

Ministry of Education and Culture
Ministère de l'Éducation et de la culture

Data rinnastetaan useissa puheissa nykyisin luonnonvaroihin, joten sille syntyy luonnollisesti myös arvo kauppatavarana. Tämän ilmiön myötä datan kerääminen ja koostaminen on synnyttänyt ja synnyttää edelleen uusia toimijoita markkinoille. Näiden datakauppiaiden toiminta puolestaan esiintyy joko yksittäisinä palveluina tai koostettuna datan markkinapaikoille. Kysyntä ja tarjonta määrittävät ennen pitkää erilaisen datan arvon, aivan kuten muillakin luonnonvaroilla. Nykyaikainen, datatietoinen organisaatio huomioi myös nämä kaupalliset datavarannot arvioidessaan liiketoimintansa datavetoista kehitystä.

Jotta big data-vallankumouksesta saadaan enemmän hyötyä irti, täytyy kumouksen raaka-aineen eli datan itsensä, olla mahdollisimman laajasti avointa ja helposti käytettävissä. Erilaisten tietoaaineistojen avaaminen ja niiden yhdistely mahdollistaa uusia palveluinnovaatioita. Avaaja ei useimmiten voi tietää, miten avattua tietoaaineistoa käytetään ja mihin sitä mahdollisesti yhdistellään. Esimerkiksi datamassan tuomaa mahdollista liiketoiminnallista potentiaalia ei voi siis lähtökohtaisesti tietää etukäteen.

Sekä EU-tason että kansallinen lainsäädäntö ohjaa erityisesti julkishallinnon dataa voimakkaasti kohti avoimuutta. Tulevaisuudessa kansallisin varoin kerättyä, varsin arvokasta dataa on paljon saatavilla ja hyödynnettävissä edelleen liiketoiminnan kehityksessä. Avoimen datan kulttuurin kehitys vetää mukanaan myös yksityisiä ja kaupallisia toimijoita, joiden intresseissä on avata dataansa julkisuus- ja goodwill-tarkoituksissa sekä vauhdittaakseen oman erityisosaamisalueensa kehitystä. Organisaatioiden onkin syytä ottaa mukaan kokonaisvaltaiseen data-strategiaansa myös avoin data ja sen erilaiset mahdollisuudet. Hyötyjä on tunnistettavissa sekä avoimen datan kuluttajana että tuottajana.

Julkisen tiedon tulee olla mahdollisimman laajasti avointa ja maksutonta. Julkisen tiedon hyödyntäminen on kirjattu myös nykyisen hallituksen hallitusohjelmaan tavoitteena. Tietovarantoja avataankin tällä hetkellä osana valtiovarainministeriön vetämää avoimen tiedon ohjelmaa.

Nykyaikainen suhtautuminen datan avaamiseen, lainsäädännön kehitykseen reagointinopeus ja tiedon jakamisen kulttuurin kehitys luovat myös kansallista pääomaa ja auttavat Suomea toimimaan esimerkkimaana. Kokemukset avoimen datan päätöksistä, käytännön toteutuksista ja vaikutuksista liiketoimintakenttään ovat lähitulevaisuudessa haluttua pääomaa. Parhaiden käytäntöjen tuotteistaminen voi luoda myös hyvää tukea suomalaisen osaamisen vientiin ja vauhdittaa osaltaan talouskasvua. Tietoon perustuva päätöksenteko vaatii tuekseen toimivan tiedonkeruun ja analysoinnin. Sama tiedonkeruu ja -analysointi palvelee myös open data -aloitteita. Avoimen tiedon tulee olla muodossa, jossa sen hyödyntäminen on mahdollisimman helppoa. Esimerkiksi erilaisten julkisten tietoaaineistojen koneluettavia standardimuotoja tulee edistää.

Datan käytettävyyteen vaikuttavat myös sen hallintamenetelmät. Eri luonteisen tiedon hallinta on tarvitsee erilaisia käytäntöjä. Tämä on äärimmäisen tärkeätä yksilöiden kohdalla. Henkilötietojen ja tiedon keräämisen ja käytön periaatteet tulisi olla hyvin tiedossa ja yleisesti luotettuja.

Esimerkiksi yksityisyyden ja käyttäjästä kerättävän tiedon osalta käytössä on erilaisia tiedon teknologisia työkaluja sekä hallinta- ja käyttömenetelmiä. Erilaiset menetelmät ja välineet sopivat erilaisiin tarkoituksiin, niillä on omat etunsa ja haittansa.^{xiii} Erilaisia menetelmiä ja välineitä ovat muun muassa:

- salaus- ja kryptausmenetelmät,
- ehtojen ilmoittaminen ja suostumuksen pyytäminen,
- tietoaaineistojen anonymisointi ja de-identifikaatio,
- tietojen tuhoaminen käytön jälkeen (deletion) tai säilyttämättömyys (non-retention)

Erilaiset menetelmät myös kehittyvät jatkuvasti. Toisaalta myös erilaiset teknologiat, joilla potentiaalisesti voidaan eri lähteitä yhdistelemällä voidaan saada selville esimerkiksi anonymisoiduista tietoaaineistoista henkilötietoja. Tätä kehityksen tasapainoa onkin syytä seurata tarkasti.

Valtiovarainministeriön vetämä avoimen tiedon ohjelma avaa julkisia tietoaaineistoja
http://www.vm.fi/vm/fi/05_hankkeet/02381_avoin_tieto/index.jsp

2.8 Lainsäädäntö, sääntely ja tietosuoja

Yksityiselämän suoja on turvattu Suomessa **perusoikeutena (PL 10 §)**. Myös Euroopan ihmisoikeussopimus ja EU:n perusoikeuskirja edellyttävät, että jokaisen oikeutta yksityiselämään on kunnioitettava. Yksityisyyden suojaa joudutaan usein punnitsemaan suhteessa muihin perus- ja ihmisoikeuksiin, kuten sananvapauteen ja julkisuusperiaatteeseen. Viime aikoina kansainvälisessä keskustelussa on myös yhä useammin noussut esiin digitaalisten perusoikeuksien turvaaminen. Suomen perustuslain mukaan henkilötietojen suojasta on säädettävä tarkemmin lailla. **Henkilötietolaki** (523/1999) on henkilötietojen käsittelyä koskeva yleislaki. Lain tarkoituksena on muun muassa toteuttaa yksityiselämän suojaa ja edistää hyvän tietojenkäsittelytavan kehittämistä ja noudattamista. Henkilötietolain velvoitteet pitää huomioida aina, kun käsitellään henkilötietoja, eli tietoja, jotka ovat yhdistettävissä tiettyyn henkilöön. Lakia ei kuitenkaan sovelleta tilanteisiin, joissa luonnollinen henkilö käsittelee henkilötietoja yksinomaan henkilökohtaisiin tai niihin verrattaviin tavanomaisiin yksityisiin tarkoituksiinsa.

Big Datan käsittelyn yhteensovittaminen henkilötietolain kanssa ei ole täysin ongelmaton. Erityisesti henkilötietolaissa kuvattu käyttötarkoitussidonnaisuus, informointivelvollisuus ja rekisteröidyn suostumus ovat Big Datan näkökulmasta haasteellisia. Tästä syystä tietojen käsittelijän on syytä varmistaa, että käsittely on henkilötietolain mukaista, jos osana Big Dataa käsitellään henkilötietoja, joita ei ole täydellisesti anonymisoitu. Esimerkiksi pelkästään nimien ja muiden yksilöintitietojen poistaminen ei vielä välttämättä tarkoita, että tiedot muuttuvat anonyymeiksi, jos henkilöt ovat edelleen välillisesti tunnistettavissa. Euroopan unionissa **valmisteilla olevan tietosuoja-asetuksen** on tarkoitus uudistaa ja harmonisoida unionin tietosuojalainsäädäntöä. Jos asetus tulee voimaan, se muuttaa monella tavoin jäsenmaiden kansallista henkilötietojen käsittelyä koskevaa sääntelyä. Big Datan näkökulmasta merkityksellisiä ovat muun muassa profilointia, tietojen käsittelyn läpinäkyvyyttä, rekisteröidyn informointia tai rekisteröidyn oikeutta tulla unohdetuksi koskevat artiklaehdotukset. Tietosuoja-asetus vaikuttaisi voimaan tullessaan kaiken sellaisen datan käsittelyyn, jota ei ole täysin anonymisoitu.

Viranomaisen hallussa olevien asiakirjojen julkisuuteen sovelletaan **julkisuuslakia (621/1999)**. Julkisuuslaki soveltuu myös silloin, kun luovutetaan henkilötietoja viranomaisen henkilörekisteristä. Henkilötietojen luovutuksen edellytyksenä on tällöinkin, että luovutuksen saajalla on henkilötietolain mukainen oikeus käsitellä kyseisiä henkilötietoja. Näin ollen julkistakaan henkilötietoa ei voida automaattisesti saattaa vapaasti uudelleen käytettäväksi, vaan luovutuksen reunaehdot määräytyvät julkisuuslain, henkilötietolain ja mahdollisen erityislainsäädännön mukaisesti. Tämä tulee huomioida myös Big Datan käsittelyssä jos käsitellään viranomaislähtöistä tietoa.

Henkilötietojen suoja ja julkisuutta koskevan sääntelyn lisäksi Big Datan käsittelyyn voi vaikuttaa **immateriaalioikeudellinen sääntely**. Erilaiset tietokannat ja niihin liittyvät ohjelmistot saattavat esimerkiksi nauttia tekijänoikeudellista suojaa. Lisäksi tietoihin ja niiden käyttöön saattaa liittyä sopimusoikeudellisia ja muita kaupallisia velvoitteita, jotka tietojen käsittelijän tulee huomioida.

Tulevaisuudessa Big Datan hyödyntäminen saattaa johtaa siihen, että sen aiheuttamia ongelmia joudutaan ratkaisemaan uuden sääntelyn kautta. Esimerkiksi syrjivät hinnoittelukäytännöt saattaisivat johtaa lainsäädännöllisiin uudistuksiin. Yritysten ja muiden organisaatioiden kannalta hyvä perusneuvo tietosuojan toteuttamiseen on, että tietojenkäsittely ensin suunnitellaan huolellisesti ja sen jälkeen pysytään tässä suunnitelmassa. Läpinäkyvyys tietojenkäsittelyssä korostuu tulevaisuudessa, mikä vaikuttaa suoraan myös Big Dataan. Kaiken kaikkiaan sekä kansallista että kansainvälistä sääntelyä tulisi kehittää siihen suuntaan, että se mahdollistaa Big Datan hyödyntämisen tarkoituksenmukaisella tavalla.

2.9 Yhteistyö ja datan vaihto

Dataintensiivisen liiketoiminta kasvaa jatkuvasti. Data on yhä arvokkaampaa: toisaalta kaikki data ei ole kaikille yhtä arvokasta. Toiselle organisaatiolle jonkin muun tahon tuottama ja tallentama, mutta ei käyttämä, data voisi olla kullannarvoista. Toimivan datanvaihdon ja datamarkkinoiden luonti on tärkeää, jotta datamoodiin siirtyvät organisaatiot saavat potentiaaliset hyödyt irti big data-vallankumoksesta.

Toisaalta muutoksen ajureita ovat olleet myös tietoyhteiskunnasta verkostoyhteiskuntaan siirtyminen, jolloin yritysten tiedon hyödyntämisen tarve on lisääntynyt merkittävästi. Jos aiemmin yritysten kilpailuetuna pidettiin suljettuja organisaatioita, jotka pyrkivät suojelemaan kaikkea hankkimaansa tietoa, uudempien liiketoimintamallien etuina on nähty siirtymistä avoimempien, yhteistyötä tekevien organisaatioiden suuntaan^{xiv}. Kilpailukykyisinä pysyvät yritykset ovat riippuvaisia myös yhteistyöverkostoista, alihankkijoistaan ja tiedon jakamisesta^{xv}. Toisaalta on tehty myös huomioita siitä, miten merkittävä osuus uusista innovaatioista tulee olemassa olevan kilpailun ulkopuolelta^{xvi}.

Eri yritysten ja organisaatioiden keräämän ja tarjoaman datan hankkimisen ja käyttämisen täytyy olla mahdollista. Tavoitteena on saada datan vaihdosta ja yhteisestä (avoimesta) käytöstä yleistä toimintaa. Datan hankkimisesta on usein sen kerääjälle kustanuksia ja datan jakamiseen omasta organisaatiosta täytyy löytyä kannusteita.

Laajempien datamarkkinoiden syntyminen vauhdittaisi huomattavasti uutta big dataan keskittyvää liiketoimintaa.

Tähän liittyy vielä kosolti haasteita. Yritysten (ja hallinnon) välinen yhteistyö on tärkeässä roolissa. Esimerkiksi pienempien innovatiivisten yritysten ja paljon dataa keräävien suurten yritysten yhteistyö on suhteellisen harvinaista. Käytännössä suurempi yritys voi muun muassa avata dataansa ja platformin kehitykselle. Toisaalta tälle pitää löytyä riittävät kannusteet myös suuremman yrityksen puolelta. Yksi mahdollisuus on selvittää, onko syntynymässä standardeja siihen, että yritykset voivat avata "sandboxeja" muiden käyttöön ja analyysiä varten.

Big datan prosessointi on erityisen haasteellista. Dataa saattaa olla niin paljon ja sitä voi syntyä niin suurella nopeudella, että sitä ei voi siirtää fyysisesti toisen organisaation käsiteltäväksi. Tällaisissa tilanteissa tiedon analysointia varten täytyy siirtää prosessoivaa koodia toisen organisaation sisälle. Tämä on ongelma, joka vaatii sekä teknistä että tietoturvallista ratkaisua. Samalla täytyy varmistua siitä, kuinka toinen osapuoli voi luottaa ulkopuoliseen ohjelmakoodiin. Yksi ratkaisu tähän voisi olla yhteiskunnan luoma luotettu dataoperaattori, joka huolehtii datan analysointikoodista ja resurssoinnista yhteistyössä osapuolten kanssa.

Big datan prosessoinnin avuksi voidaan valjastaa myös kiinteiden organisaatioiden ulkopuolista prosessointivoimaa muun muassa joukkoistamalla ja avaamalla tietovarantoja prosessointiin ja innovointiin avoimiin tietoverkkoihin. Jos verkostoituneen todellisuuden mahdollisuuksia ei osata hyödyntää täysimittaisesti, saattaa mittaamattoman arvokkaiksi koettuja tietopääomia panttaavat yritykset näivettyä vanhan maailman mummuteiksi. Kehittämispotentiaalia ja luovuutta on kaikissa yritysten sidosryhmissä ja verkostoissa. Edellä mainitusta on jo useita vuosia puhuttu uudesta taloudesta, Wikinomicsista, jossa yritysten johtajat saattavat tiukasti vanhanaikaisista liiketoimintamalleistaan ja tiedostaan kiinni pitämällä tuomita yrityksensä joko

Saksassa yhteistyötä tehdään muun muassa suuret ja pienet yritykset sekä tutkimuspuolen yhteen tuovan Smart Data Innovation Labin kautta
http://www.kit.edu/kit/english/pi_2014_14408.php
<http://www.sdil.de/de/>

Suomessa Digilen vetämä D21-ohjelma kokoaa yhteen suuria yrityksiä, pk-yrityksiä ja tutkimuspuolta löytämään yhdessä ratkaisuja big datan hyödyntämiseen.
<http://www.datatointelligence.fi/>

menestymään tai tuhoutumaan. Tapscottin ja Williamsin mukaan avoimuus niin tiedollisesti kuin asenteellisesti, yhteistoimintafoorumeilla tapahtuva jakaminen, ja globaalisti toimiminen, niin alueellisesti kuin organisaatorajojenkin ulkopuolella, ovat tulevien menestystarinoiden reseptejä.^{xvii}

Suomen laaja start-up-kenttä (esim. AaltoES) tulee kytkeä tiiviisti mukaan big data-ratkaisujen ja palvelujen kehitykseen. Yritysten ja ketterien start-uppien kokeilevaa yhteistyötä tulee lisätä.

Yhteistyö voi olla myös kansainvälistä. Suomalaisten yritysten, esimerkiksi pk-sektorilta, sekä kansainvälisten veturiyritysten yhteistyötä tulisi aktiivisesti edistää. Esimerkiksi Saksa on avainasemassa teollisen internetin tulevissa ratkaisuissa. Suomalaisilla yrityksillä ja niiden osaamisella on paljon erilaisia mahdollisuuksia näiden globaalisti johtavien yritysten big data-kehityksessä.

2.10 Kokeilut ja t&k rahoituksen suuntaaminen

Big data-kokeiluille tulisi saada riittäviä resursseja ja rahoitusta. T&K-rahaa tulisi kohdentaa big dataan liittyville kokeiluille ja tutkimukselle. Lisäksi muun muassa EU:n Horisontti 2020-ohjelma tarjoaa hyviä mahdollisuuksia.

Haastetta ketterien kokeilujen toteuttamiselle asettaa se, että suurten tietovarantojen haltijat ja niitä kerryttävät tahot ovat usein suuria eivätkä aina niin ketteriä organisaatioita. Toisaalta data-analytiikan alueelle on syntynyt ja syntymässä uusia pieniä yrityksiä, jotka voisivat tarjota tehokkaampia menetelmiä big datan hyödyntämiseen ja uuden liiketoiminnan synnyttämiseen. Myös T&K&I -rahoitusta suuntaamalla olisi mahdollista saattaa erilaisia big-data -toimijoita yhteisiin projekteihin ja hakemaan yhdessä ratkaisuja todellisiin asiakastarpeisiin.

Ylipäänsä big dataan liittyviä ja kohdentuvia tutkimusinvestointeja täytyy tehdä, mikäli Suomi pyrkii pysymään kehityksessä mukana. Big data-kehitykselle ollaan muissa maissa osoitettu huomattavia investointeja valtion taholta.

Suomessa julkisten toimijoiden tulee osittain paikata yksityisten pääomasijoitusten puutetta. Esimerkiksi Tekesin eri ohjelmien kautta saatavaa rahoitusta tulee kohdistaa big data-kehitykseen. Oleellista on saada myös yritykset mahdollisuuksiensa mukaan lisäämään riskirahoitustaan big dataan liittyvien.

EU:n Horisontti 2020-ohjelman toisessa pilarissa, ”johtoasema mahdollistavissa teollisuusteknologioissa”, big datalle on omistettu kaksi rahoitusohjelmaa. Toinen liittyy big datan ja avoimen datan innovaatioiden levittämiseen ja toinen big data-tutkimukseen.

<http://ec.europa.eu/programmes/horizon2020/en/h2020-section/information-and-communication-technologies>

2.11 Oman tiedon hallinta

Kansalaisten tulee voida vaikuttaa siihen, mihin ja miten heidän tietojaan käytetään. Kerättyä tietoa tulisi halutessaan voida hyödyntää myös eteenpäin.

Tiettyjä tietoja luovuttamalla voisi saada esimerkiksi parempaa ja yksilöllisempää tai halvempia palveluita. Samalla tämä saattaisi mahdollistaa uudenlaisten palveluiden kehittämisen. Useimmiten jo nyt pyydettyä luovutettavien tietojen laajempi hyödynnettävyys olisi myös tietoja keräävien yritysten intresseissä. Niin sanottu Mydata-kehitys voisi tarjota yhden mahdollisen tavan omien tietojen hyödyntämiseen ja hallitsemiseen. Ratkaistaviin kysymyksiin kuuluu muun muassa, missä tietoa säilytetään sekä missä muodossa tietoa tulisi luovuttaa. Samalla yksittäisen kansalaisen (asiakkaan)

LVM teettää selvitystä Mydatasta. Selvityksen tavoitteena on kiteyttää tulevaisuuden skenaariot henkilökohtaisen datan hallinnoinnin, yhdistämisen ja jakamisen alueella sekä luoda pohjaa suomalaisen henkilökohtaiseen dataan perustuvan yhteistyöverkoston syntymiselle. Selvitys valmistuu 16.9.2014 mennessä

<http://www.lvm.fi/web/hanke/liikenteen-ja-viestinnan-avoin-tieto>

kannalta omien tietojen hallinta ja hyödyntäminen mahdollistaa mm. säästöjä, juuri itsensä kannalta sopivien palveluiden "räätälöinnin" sekä luottamuksen lisäämisen tietoja kerääviä yrityksiä kohtaan.

3. Sovellusalueet ja niiden potentiaali Suomessa

Big datan käyttö-työryhmä on tunnistanut ja priorisoinut Suomen kannalta oleellisia sovellusalueita. Osassa Suomella on laajaa osaamista ja jopa käyttöä. Osa sovellusalueista on taas talouden ja koko yhteiskunnan kannalta niin oleellisia, että erilaisia big data-ratkaisuja tulisi saada mahdollisimman pian käyttöön, jotta Suomi pysyy mukana muun muassa rajussa kansainvälisessä kilpailussa.

3.1 Terveys

Terveysalalla Suomella on suuri potentiaali big datan suhteen. Suomesta löytyy maailmanlaajuisesti katsoen poikkeuksellisen laadukkaita ja kattavia tietokantoja, muun muassa geenitietoihin liittyen. Näiden vahvuuksien vastuullinen käyttäminen big data-ratkaisuissa voisi tarjota Suomelle merkittävän, globaalin kilpailukykytekijän.

Terveystietojen kenttä liittyy big dataan lukuisilla eri tavoilla, ja Suomessa on jo alan osaamista esimerkiksi bioinformatiikan ja molekkyylilääketieteen alalla. Toinen tärkeä linkitys tulee itsenmittausteknologioiden kehityksestä, jossa suomalaisyrityksissä on osaamista kymmenien vuosien ajalta. Terveysdatan laajemman saatavuuden ennustetaan muuttavan ihmisten käsitystä itsestään ja omasta terveydestään. Elämäntavoista ja geenitiedon avulla kertyvän big datan avulla voi ehkä tulevaisuudessa ehkäistä sairastumista. Nykyään terveystietoja voi kerätä esimerkiksi Taltioniin, josta voi tulla tulevaisuudessa tärkeä alusta henkilökohtaiselle big datalle.

Big data mahdollistaa terveydenhuoltoon uudenlaisen oppimisen, kun kerrytetty ja analysoitu data tukee lääkärin ja muun hoitohenkilökunnan työtä. IBM:n kehittämä Watson on yksi esimerkki siitä millaisia käyttöjä big datalle on kehitelty. Kun potilas kertoo oireistaan lääkärille, lääkäri voi samalla kysellä Watsonilta taustatietoja potilaan oireiden perusteella. Watson käy muutamassa sekunnussa tietokannastaan läpi hakusanojen avulla tuhansia oppikirjoja ja satoja tuhansia tieteellisiä artikkeleita.

Big datalta odotetaan terveydenhuollon laadun parantumista ja säästöjä. Data-analyysin odotetaan kehittyvän tavalla, joka auttaa yksittäisen kansalaisen terveystietojen ennustamisessa erilaisten datavirtojen avulla sekä sairauksien ehkäisyssä. On tärkeää, että suomalaiset tuntevat kansainvälisen kentän ja ovat mukana ohjaamassa ajankohtaisia kehityskulkuja.

Terveystietojen hyötykäyttöön liittyy yksityisyyteen, sosiaalisiin ja eettisiin näkökulmiin liittyviä kysymyksiä, jotka on otettava vakavasti. Ilman niitä suotuisat kehityskulut tuskin ovat mahdollisia. Yksi tapa on määritellä vastuullisen aineiston käytön puitteita yhteistyössä eri alojen osaajien kanssa.

Suomessa on muun muassa Sote-uudistuksen myötä oiva tilaisuus pohtia tarkemmin, kokeilla ja ottaa käyttöön big dataan perustuvia ratkaisuja. Dataan perustuvista hoitomenetelmistä ja -käytännöistä on saatu merkittäviä tuloksia. Big dataan perustuvien ratkaisuilla voidaan saavuttaa selviä taloudellisia säästöjä. Nykyisessä taloudellisessa tilanteessa, varsinkin kuntien osalta, säästöjen saavuttaminen uuden teknologian avulla palvelutasoa sinänsä laskematta on avainasemassa. Erilaisten dataan perustuvien omahoitomenetelmien käyttöönottoa tulee samalla tukea.

3.2 Älykkäät verkot

Datan merkityksen lisääntyminen heijastuu myös yhdyskuntasuunnitteluun ja energianjakeluun. Kaupungit kehittyvät ns. "smart city" -ajattelun mukaisiksi, joissa älyverkkojen mahdollistamat uudet energianjakelun ja asumisen ratkaisut tuodaan osaksi arkipäivää. Älyinfran rakenteet huomioidaan kaavoituksessa, ja suomalaista smart grid-osaamista voidaan käyttää täysipainoisesti hyödyksi. Nämä älykkäät verkot tarvitsevat toimiakseen toimivan verkkoinfran lisäksi myös älysovelluksia, joissa hyödynnetään avoimia tietomassoja ja olemassa olevia tietovarantojen infrastruktuureja. Näin ollen voidaan puhua älyverkko-osaamisesta, johon oleellisesti linkittyy big data sekä Suomen ICT-osaaminen.

Suomessa Smart City -ajattelun kehityskärjessä kulkevat useat tutut yritykset, kuten IBM, Siemens, ABB, Schneider, sekä Helen ja Fortum. Uusia palvelumalleja kehitetään jatkuvasti sellaisiksi, että ne pystyvät hyödyntämään parhaalla mahdollisella tavalla uusia liiketoiminnan ekosysteemejä. Asumisen innovaatioissa tällaisia palveluratkaisuja ovat mm. kotiin asennettavat anturit, jotka mittaavat sisäilman laatua. Smart Cityjen ja älykkäiden verkkojen kasvualustana ovat big data, valtavasti kasvanut laskentateho, anturiteknologian kehittyminen, sekä automatisoinnin mukanaan tuomat mahdollisuudet.

Energiatehokkuuden lisäksi myös resurssien käyttö on Smart Cities -ajattelussa entistäkin tehokkaampaa, ja synnyttää ympärilleen uusia älypalveluita. Älykkäiden verkkojen avulla älykaupunkien älytalot tuottavat valtavat määrät dataa, jota voidaan käyttää uusien Smart City -palvelujen rakennusaineena.^{xviii} Smart grid -sähköverkot mahdollistavat sähkön myymisen takaisin verkkoon, jolloin erilaisia sähkölaitteita on kytketty älykkäästi toisiinsa, säilömään ja varastoimaan ylimääräistä energiaa. Älyverkkojen merkitys on erityisen korostunut kylmissä olosuhteissa, jolloin älysensorit voivat sammuttaa käyttämättömät osat rakennuksista energiatehokkuuden optimoimiseksi tai säätää lämpötilaa tarpeen mukaan. Tämä tasaa myös kulutuspiikkejä ja lisää dynaamisuutta, sekä lisää mahdollisuuksia säätää sähkönkulutus tapahtumaan aikoina, jolloin uusiutuvaa energiaa on saatavilla.

Älykkäiden sähköverkkojen MDMS-hallintajärjestelmällä (Meter Data Management System) älydata kerätään ja hallinnoidaan. Sensoreista ja mittareista saatava tieto voidaan ohjata myös uusien palvelujen ja sovellusten kehittämiseen. On ennustettu, että mitattavan datan hallintajärjestelmien markkinat kasvavat yli 300 % vuonna 2014^{xix}.

Mobiilidatatilastot kaupunkisuunnittelun apuna

Aiemmin staattisten asutustilastojen kautta tehtyä kaupunki- ja tilasuunnittelua voidaan tehostaa merkittävästi käyttämällä avuksi matkaviestinten verkkodatatilastoja. Dynaamiset mallit ovat tärkeässä asemassa, sillä sekä ikä- että palvelurakenteiden muutokset sekä väestön liikkeiden arviointi tarvitsee yhä täsmällisempiä ennusteita ja malleja. Näistä malleista on apua mm. tie- ja liikennesuunnittelussa, maankäytön suunnittelussa, maisemoinnissa sekä puisto- ja viheraluesuunnittelussa.

Japanin suurin mobiilipalvelujen tarjoaja NTT Docomo teki yhteistyössä Tokion yliopiston kanssa tutkimuksen yliopiston Kashiwan kaupungin kaupunkisuunnittelun avuksi. Ensimmäisessä tapaustutkimuksessa saatiin tietoa maankäytöstä ja asukkaiden liikkeistä ryhmänä, kun alue jaettiin kolmeen alueeseen (keskusta, asuinalueet ja maaseutualueet). Tämä tieto osoittautui kaupunkisuunnittelun kannalta hyvin oleelliseksi. Toisessa tutkimuksessa selvitettiin asukkaiden puistojen ja viheralueiden käyttöä, ja tästä tutkimuksesta saatua tietoa voitiin käyttää perustietoa puistojen sijoittamista koskevana perustana.

Kolmannen tutkimuksen merkittävä anti oli julkisen liikenteen tarpeiden mitoittaminen ihmisten asuinalueiden mukaan. Jakamalla asukkaat asuinalueiden ja vuorokauden aikojen mukaan ryhmiin saatiin tietoa siitä, mitkä ihmisten liikkumistarpeet ovat kunakin vuorokauden aikana. Vertaamalla tätä tietoa paikallisliikenteen tietoihin voitiin saada tietoa julkisen liikenteen palvelujen helppokäyttöisyyden lisäämiseksi. Neljännessä tutkimuksessa kaupungin keskustassa käyneet ihmiset ryhmiteltiin yksityiskohtaisesti kellonajan, vuorokauden, iän ja sukupuolen mukaan, jolloin tätä tietoa voitaisiin käyttää paikallisen tiedotuksen tarpeisiin.

Suurin tutkimusten anti koski ihmisten liikkeitä tietyllä alueella palvelujen tehostamiseksi ja parantamiseksi. Tutkimusryhmän mukaan lisätietoa voidaan tulevaisuudessa saada, kun teletunnistetiedoista saatavia tietoja yhdistellään muihin saatavilla oleviin tietoaaineistoihin.

Using Mobile Spatial Statistics in Field of Urban Planning. Odawara, T. & Kawakami, H. 2013. NTT Docomo Technical Journal Vol. 14 No. 3, p. 31-36.

https://www.nttdocomo.co.jp/english/binary/pdf/corporate/technology/rd/technical_journal/bn/vol14_3/vol14_3_031en.pdf

3.3 Tutkimus

Big data-kehitys tarjoaa tutkimukselle huomattavia mahdollisuuksia. Esimerkiksi uusien, dataan perustuvien tutkimusmenetelmien kehitys tuo lähes alan kuin alan tutkijoille huomattavasti parempia mahdollisuuksia tutkia erilaisia asioita ja toisaalta löytää ongelmiin myös vastauksia. Muun muassa erilaisen mobiilidatan, trendidatan sekä sosiaalisesta mediasta saatavan datan määrän kasvu tuo tutkijoiden ulottuville aivan uudenlaisia aineistoja.

Big dataa voidaan hyödyntää monilla tutkimusaloilla. Esimerkiksi biotieteiden, tähtitieteen, fysiikan ja vaikkapa psykologian (psykonometria) ja taloustieteen (ekonometria) alalla big data-kehityksestä ja menetelmistä on saatavilla selkeitä hyötyjä. Muun muassa näiden alojen tutkimukseen big data tarjoaa täysi uusia mahdollisuuksia

Menetelmäkehityksen lisäksi tärkeää on muistaa niiden hyödyntämisessä muun muassa monitieteisyys. Esimerkiksi erilaisiin yhteiskunnallisiin ongelmiin voidaan hakea vastauksia uusien menetelmien avulla. Tämä vaatii eri tieteenalat ylittävää työskentelyä, muun muassa matemaatikkojen, ohjelmoijien ja yhteiskuntatieteilijöiden kesken. Erilaiset laitos- ja tieteenalarajat eivät saa estää uusien menetelmien sekä mahdollisuuksien täysimääräistä käyttöä. Yliopistojen tulisi olla valmiita muun muassa erilaisten monialaisten tutkimustiimien tukemiseen.

3.4 Liikenne

Liikenteen alalla datan räjähdysmäinen lisääntyminen ja big data-ratkaisut tarjoavat selkeitä mahdollisuuksia. Erilaisilla ratkaisuilla tavoitellaan muun muassa liikenneturvallisuuden parantumista ja liikenteen tehostumista sekä liikenteen ekologisuutta. Ratkaisut voivat liittyä muun muassa liikenteen ohjausjärjestelmiin, logistiikan optimointiin sekä erilaisiin reittipalveluihin.

Suomella on mahdollisuus profiloitua liikenteen kokeilu ympäristönä^{xx}. Tätä puoltaa muun muassa Suomen suhteellisen pienimuotoinen liikenne, jolloin pienempiä kokeiluja voidaan hyvin tehdä. Suomessa tehdään myös alan tutkimusta ja ylipäänsä ollaan maailmassa kärkijoukossa esimerkiksi avoimen liikennedatan käytön suhteen. Suomessa on myös tehty tutkimusta uudenlaisista liikennejärjestelmistä.

Parhaimmillaan erilaiset liikenteeseen liittyvät big data-innovaatiot (esim. liikenneanalyysi ja -ennusteet) voivat olla merkittäviä vientituotteita. Erilaista tutkimus- ja innovaatiotukea sekä muuta rahoitusta on liikenteen alalla saatavissa suhteellisen runsaasti. Näin ollen mahdollisuus merkittävään toimintaan on olemassa.

3.5 Teollinen internet

Huolto ja ennakoivat etäpalvelut ovat yksi osa-alue laajemmassa kokonaisuudessa, josta usein käytetään termiä Teollinen Internet (TI). Tällä tarkoitetaan sulautettujen ja älykkäiden laitteiden ja järjestelmien, niistä jatkuvasti kertyvän datan ja siihen pohjautuvan data-analytiikan sekä ihmisten työn tehokasta yhdistämistä liiketoimintaprosesseissa. Tämä mahdollistaa sen, että tuotanto- ja muut resurssit, tieto, esineet ja ihmiset muodostavat reaaliaikaisesti verkottuneen kokonaisuuden.

GE:n valmistamista lentokoneturbiineista kerätään päivässä huimia datamääriä. Yksi turbiinin sensori kerää päivässä noin 500 gigaa dataa. Yhdessä turbiinissa taas on 20 sensoria ja GE:llä on noin 12 000 turbiinia.

<http://sites.tcs.com/big-data-study/ge-big-data-case-study/>

<http://www.lopezresearch.com/2013/06/04/the-new-industrial-revolution-according-to-ge-and-ptc/>

Myös suomalaiset teollisuusyritykset käyttävät big data-ratkaisuja muun muassa ennakoivaan huoltoon. Hyviä esimerkkiyrityksiä ovat Konecranes, Wärtsilä ja KONE. <http://www.bigdata.fi/artikkelit/big-data-ennakoiva-huolto>

Sovellusalueita ovat mm. valmistavan teollisuuden prosessit ja niiden optimointi, ennakoiva huolto, energian käytön hallinta, käyttöomaisuuden hallinta ja ennakoiva huolto. Vähintään yhtä suuret hyödyntämismahdollisuudet ovat myös varsinaisen teollisuuden ulkopuolisessa elinkeinolämässä, kuten esim. terveydenhuollossa, kaupan ja logistiikan alueella,

rakentamisessa ja kiinteistöjen hoidossa sekä kunnallisten ja muiden julkisten palvelujen tuottamisessa (energia, vesi, jätevesi...). Yhteistä kaikille sovellusalueille on se, että tavalla tai toisella on olemassa automatisoitu linkki fyysisen maailman ja digitaalisen maailman välillä. Usein tuo linkki syntyy joukosta antureita tai tägejä, joilla voidaan saada tietoa esineiden tai ihmisten tilasta, olinpaikasta ja muista tekijöistä. Kun tätä tietoa yhdistetään ja analysoidaan yhdessä kertyneen historiatiedon sekä muiden tietovarantojen kanssa, voidaan tehostaa merkittävästi nykyisiä toimintatapoja ja -prosesseja sekä luoda myös aivan uusia palveluja ja liiketoimintaa. Alan uusissa palveluissa on myös huomattava vientipotentiaali. Toisaalta esimerkiksi etähuoltoratkaisujen avulla voidaan pitää ja saada lisää korkean jalostusarvon työpaikkoja Suomeen.

Suomella olisi mahdollisuuksia muun muassa palveluliiketoiminnan synnyttämisessä teollisen internetin avustuksella ja -ympäristöille. Palvelu ja huolto yhdistettynä etähuoltoon, automatisointiin, ennakoivaan huoltoon. Merkittävä osa ongelmanratkaisusta ja korjauksista voidaan automatisoida. Tämä toiminta sopii eri teollisuusalueille: Metalliteollisuus, konepajat (esim. Konecranes), tietoliikenne. Kaikki vähänkin arvokkaammat laitteet tai toiminnallisuudet pystyvät itse analysointiin ja ne voidaan testata/analysoida/korjata etäältä. Usein vikaantumisen voidaan jo ennakoita. Tällä toiminnan optimoinnilla saadaan merkittäviä säästöjä ja kompetenssi siirretään halpatyösuunnasta korkeamman kompetenssin vaatimuksiin.

Eri toimialoilla samoja tai samantapaisia tarpeita. Teollisen internetin kehityksen hyödyntäminen vaatii osaamista ja kombinaatiota erilaisista asioista. Näihin lukeutuvat muun muassa kompetenssi, anturien kehitys, etäyhteyksien parantuminen sekä Big Data analytiikka. Haasteena teollisen internetin kehityksessä Suomessa on muun muassa se, että kovin laajamittaista hyödyntämistä

Suomessa on alaa silmällä pitäen hyvä koulutustaso ja kompetenssi, jokseenkin osaajapula voi ainakin big data-osaajien kohdalla uhata tulevaisuudesa. Myö toimintaa ja intressi on jo osittain olemassa. Etäinen sijainti päämarkkinoilta ja päämyyntialueilta on Suomelle luonnollisista. Tämä tarjoaa tiettyssä mielessä otollisen . Sopivia teollisuusaloja, joihin big data-lähtöinen palveluliiketoiminta on sovellettavissa, ovat muun muassa metalli-, kone- sekä tietoliikenneteollisuus. Kaikilla näillä on perinteisesti ollut Suomessa tärkeä rooli viennin kannalta. Toisaalta nämä alat ovat osittain olleet kriisissä ja vaativat uudistumista. Big dataan perustuvat ratkaisut ovat avainasemassa kansainvälisessä kilpailussa mukana pysymiselle.

3.6 Cleantech

Big data liitettynä cleantech-osaamiseemme mahdollistaa uusia innovointi- ja vientimahdollisuuksia mm. jättimäisille ja kasvaville Aasian markkinoille. Suuri osa Cleantech- ratkaisuista perustuu kerääntyvän digitaalisen tiedon parempaan hyödyntämiseen ja toiminnan optimointiin tämän pohjalta. Erilaisilla big data-ratkaisuilla onkin Cleantech-kehityksen kannalta oleellinen rooli.

Cleantechin yleinen merkitys on vahvassa kasvussa resurssiniukassa maailmassa, ei välttämättä aina omana ympäristötekniikan alanaan vaan kaikkeen muuhun liiketoimintaan sulautettuna.

Myös Cleantech-kehitystä silmälläpitäen tarvitaan uusi platform-ajattelua hyödyntävä yhteistyön malli, jolla pienet yritykset voisivat suoraviivaisemmin toimia toimia isojen teollisuusyritysten innovaatiopartnereina ja sitä kautta tuottaa runsaasti ja tehokkaasti yksinkertaisia konsepteja/palveluita suuryritysten teknologian (alustat, laitteet) tai datan ympärille

Cleantech on yksi työ- ja elinkeinoministeriön nimeämistä kasvun kärjistä. Valtioneuvoston periaatepäätös kasvun uusien kärkien cleantechin ja biotalouden vauhdittamisesta.

https://www.tem.fi/files/39772/VNP_kasvun_karjet_cleantech_ja_biotalous_08052014.pdf

Nopean kansainvälistymisen haasteet käännettävä vahvuudeksi uusilla vientikelpoisilla innovaatioilla ja palveluilla. Cleantechin innovoinnista ja vahvuuksista huolimatta Suomella ei välttämättä juurikaan ole alueelta omia,

kansainvälisesti tunnistettuja kotimaisia käyttöreferenssejä. Ratkaisuja täytyisi siis kokeilla rohkeasti ensin myös kotimaassa. Osittain haasteena ovat myös teollisen internetin standardien ja osin teknologioiden kypsyttömyys, vaikka tulevaisuuden potentiaalia on paljon. Cleantech kulkee siis käsi kädessä teollisen internetin kehityksen kanssa.

3.7 Markkinointi ja mainonta

Gartnerin mukaan 50% Big Dataan käytettävistä investoinneista menee markkinointiin. Markkinointi on muuttumassa teknologiavetoiseksi. Gartnerin mukaan Chief Marketing Officerit käyttävät vuonna 2017 enemmän rahaa IT:hen kuin Chief Information Officerit.

Big Datan tyypillisimmät käyttötapaukset markkinoinnin ja mainonnan alueella liittyvät kuluttajakokemuksen parantamiseen, sisällön personointiin ja viestinnän kohdentamiseen.

Big Datalla on suuri rooli muun muassa seuraavissa markkinoinnin ja mainonnan aloissa:

- * Consumer Experience Management
- * Real-time Bidding
- * native advertising
- * mobile & video advertising
- * location-based advertising

4. Julkishallinto big datan sovellusalueena

Suomen hallitus on pyrkinyt torjumaan kestävyysvajetta ja tasapainottamaan julkista taloutta mm. rakennepoliittisilla uudistustoimenpiteillä, jotka kohdistuvat moniin eri yhteiskunnan osa-alueisiin kuten esim. terveydenhuoltoon, kuntauudistukseen, työurien pidentymiseen, rakenteelliseen ja nuorisotyöttömyyden torjuntaan sekä työ- ja kouluttautumismahdollisuuksien lisäämiseen. Rakenteellisten muutosten mahdollistamien kustannussäästöjen lisäksi tuottavuuden lisääminen niin kustannus- kuin palvelutehokkuutta parantamalla on keskeisiä julkisen sektorin tavoitteita.

Big data ajattelutapana (kts. datatietoisuus) ja teknologiana antaa omalta osaltaan uudenlaisia näkökulmia hallinnolle edistää tuottavamman yhteiskunnan ja kestävyysvajeen torjumisen strategisia päätavoitteita, lisäten samalla kansalaisten tyytyväisyyttä julkisiin palveluihin. Big datan avulla on mahdollista realisoida tuottavuushyötyjä useimmilla hallinnon alueilla. Datalähtöisempää julkishallintoa voidaan yleisesti tarkastella esim. kolmessa seuraavassa osa-alueessa: Datalähtöisen päätöksenteon ja jatkuvan organisaatiokehityksen tuominen osaksi julkishallinnon kulttuuria, kansalaisten henkilökohtaisemmat palvelut sekä yritysten ja kansalaisten parempi osallistaminen mm. julkisten palveluiden kehitykseen (kts. Osallistaminen) Lisäksi Big datan hyödyntämistä tulee tarkastella eri hallinnonaloille kohdistettavissa olevien yksittäisten Big data-käyttötapausten kautta.

Big datan tuottavuusparannuksista julkishallintoon on haastavaa esittää tarkkoja arvioita ilman täsmällisempää analyysiä, mutta McKinsey arvioi vuoden 2011 raportissaan Euroopan julkishallintojen kykenevän Big Dataa hyödyntämällä vähentämään hallinnollisia kulujaan n. 15-20 prosenttia ja lisäämään 0,5 prosentin vuotuisen parannuksen tuottavuuteen.^{xxi} Tavoitteena data- ja asiakaslähtöisempi hallinto

Datalähtöinen organisaatiokulttuuri edellyttää, että julkishallinto pyrkii tietoisesti ja jatkuvasti omaksumaan maailmalla parhaiten toimiviksi todettuja käytäntöjä tiedolla johtamisessa, organisaatioiden kehittämisprosesseissa (esim. lean-prosessit), organisaation tavoitteisiin sidotussa suorituskykymittauksessa sekä dataan perustuvan päätöksenteossa. Datalähtöisessä julkishallinnossa organisaatiot, niin pienet kuin suuremmat yksiköt, määrittelevät omaa missiotaan ja tavoitteitaan tukevat, selkeästi ymmärrettävät ja läpinäkyvät datalähtöiset suorituskykymittarit.

Kansainvälisestikin vertailukelpoisesti valitut mittarit ohjaavat organisaatioita tekemään ohjattuja päätöksiä, jotka palvelevat mahdollisimman tehokkaasti organisaation oman mission ja kansalaisten palvelutulosten toteutumista. Datalähtöistä päätöksentekoa tukevat Big Data-järjestelmät mahdollistavat päätöksenteon tavalla, jossa tekeillä olevien päätösten hyötyjä, kustannuksia ja pitkän tähtäimen vaikutuksia on mahdollista ymmärtää ja simuloida ennen päätösten varsinaista toimeenpanemista.

Big Datalla on tärkeä merkitys entistä edistyneempien ja ennakoivempien mittareiden mahdollistamisessa, samoin kuin laadukkaiden mittarien edellyttämän hienojakoisten signaalien löytämisessä organisaatioiden suurista tietomassoista. Muutos datalähtöiseksi organisaatioksi on kokonaisvaltainen, joskin eri kypsyysvaiheessa olevat organisaatiot toki ottavat askeleita tavoitteisiin vaiheittain. Datan avaamisesta, laadun varmistamisesta, tiedonhallinnan roolien, prosessien ja vähimmäisarkkitehtuurikyvyksien luomisesta edetään tietoisesti kohti omaa toimintaansa mittaavaa ja optimoivaa organisaatiota. Lopulta oleellista on valita organisaation

Ranskan datalähtöinen hallintoreformi

Esimerkkinä julkishallinnon mission asiakaslähtöistämisestä voidaan pitää Ranskan massiivista finanssikriisin jälkeistä hallintoreformia, jonka säästötavoite vuoden 2009-2013 aikana oli 15 miljardia euroa ja leikkaustarve yli 150 000 julkisen sektorin tehtävää. Kustannusäästöjen lisäksi tavoitteena oli palvelun laadun radikaali parantaminen. Reformia valmistellessa kansalaisilta ja yrityksiltä pyydettiin yksinkertaista pisteytysarviota koskien yhteensä 50 erilaista kansalaisten ja julkishallinnon välistä "elämänaikaista tapahtumaa" sekä vastaavasti 30 erilaista yritysten ja julkishallinnon välistä "yritysten elinkaarenaikaista tapahtumaa". Esimerkkeinä kansalaisia koskevista tapahtumista olivat esim. "avioliiton solmiminen" ja yrityksiä koskevista interaktioista esim. "uuden yrityksen perustaminen". Reformin alussa kansalaiset ja yritykset arvioivat kunkin julkishallintoon suuntautuvan interaktionsa koettua monimutkaisuutta arvoilla "yksinkertainen", "OK", "monimutkainen", "erittäin monimutkainen".

Varustettuna näillä alustavilla kansalais- ja yritysnäkökulmilla julkisen asiointin havaitusta monimutkaisuudesta, tuli kyseisiä asiointiprosesseja hallinnoivien yksiköiden uudeksi ja ainoaksi asiakaslähtöiseksi missioksi uudistaa koko toimintansa siten, että kansalaisten ja yritysten monimutkaiseksi kokemat "elämäntapahtumat" saadaan mahdollisimman nopeasti muutettua mahdollisimman yksinkertaiseksi. Reformin valmistuttua kyselyn tulokset osoittivat julkisen asiointin havaitun monimutkaisuuden pudonneen (eli koetun palvelun laadun lisääntyneen) 20 prosenttia kansalaisten tapauksessa ja 25 prosenttia yritysten tapauksessa. http://www.mckinsey.com/insights/public_sector/transforming_government_in_france

suorituskykyä kuvaavat korkean tason suoriutuskykymittarit siten, että ne voidaan ymmärrettävästi ja läpinäkyvästi linkittää organisaation mission ja tavoitteiden toteutumiseen ja näitä edistäviin käytännön toimenpiteisiin. Mittareiden pohjana olevien oletusten, käsittelysääntöjen ja lopulta itse datan läpinäkyvyys kansalaisille toimii omalta omaltaan takeena valittavien mittareiden luotettavuudelle ja "peukaloimattomuudelle". Julkisten toimijoiden tärkeimmät korkean tason suorituskykymittarit kohdistuvatkin yleensä lopulta toimijoiden palvelemaan "ulkoisiin asiakkaisiin" – eli kansalaisiin ja heidän palvelukokemukseensa, mikä tilivelvoittaa hallintoa jatkuvasti parantamaan kansalaisille tarjoamaansa palvelua. Yleisiä, koko Suomea koskevia taloudellisia ja yhteiskunnallisia korkean tason mittareita ylläpitää esim. Findikaattori (www.findikaattori.fi).

4.1 Suorituskykyä parhailla käytännöillä

Julkishallinnon yksiköiden oman päätöksenteon tehostamisen lisäksi datalähtöisellä mittaamisella on erittäin tärkeä merkitys julkisten yksiköiden keskinäisessä suorituskykyvertailussa. Se tarjoaa julkisen toimijan ulospäin läpinäkyvän suorituskykyinformaation lisäksi yksikölle itselleen vahvan sisäisen insentiivin ja lähtökohdan kehittää omaa suorituskykyään edelleen. Suorituskykyvertailun keskeisenä oivalluksena on se, että näennäisesti erilliset julkiset toimijat voivat olla missioltaan, perusprosesailtaan tai esim. rakenteeltaan hyvinkin samankaltaisia.

Nämä keskinäiseen vertailuun soveltuvat yksiköt voivat olla esim. eri henkilökunnan tai tahon toimesta johdettuja sisäyksiköitä samassa julkisessa emo-organisaatiossa tai hallinnonalassa mutta aivan yhtä hyvin verrokkiorganisaatiot voivat olla kokonaan erillisiä

julkisia toimijoita tai yksityisiä yrityksiä Suomesta tai muualta maailmasta. Kun julkiset

Verokarhulle tehoa datalähtöisellä benchmarkkauksella

Esimerkkinä suorituskykyvertailusta on tutkimus, jossa yhteensä 13 maan verohallintoa verrattiin toisiinsa erilaisilla verohallinnollisilla osa-alueilla (esim. verotietoja koskevat rutiinitoimenpiteet, ei-ilmoitettuihin tietoihin liittyvät auditoinnit, tehostettu proaktiivinen veronkeruu, monikanavainen asiakaspalvelu). Tiivistäen, tutkimus osoitti että yksikään verohallinnoista ei ollut paras kaikilla osa-alueilla eli kaikilla oli opittavaa toisiltaan suorituskyvyn parantamisessa. Lisäksi tutkimus osoitti, että 13 verohallintoa yhdessä voisivat kerätä 86 miljardia euroa lisää suorina veroina ja säästää 6 miljardia kuluissa, mikäli kaikki verohallinnot toteuttaisivat veronkeruun yhtä hyvin kuin paras kolmasosa. (Thomas Dohrmann and Gary Pinshaw, "The road to improved compliance: A McKinsey benchmarking study of tax administrations 2008-2009")

http://www.mckinsey.com/client_service/public_sector/expertise/public_finance

toimijat ja yritykset eri puolilla Suomea ja maailmaa mittaavat omaa suorituskykyään lisääntyvässä määrin keskenään vertailukelpoisella standardimittaristolla, voivat tietyllä osa-alueella muihin nähden alisuorittavat julkishallinnon yksiköt oppia, omaksua tai parhaassa tapauksessa suoraan kopioida parhaita suorituskyvyn tehostamiskäytäntöjä niiltä verrokkiyksiköiltä, jotka ovat kyseisellä osa-alueella kansainvälistä huippua.

Edellä esitetyt datalähtöisemmän organisaatio- ja päätöksentekokulttuurin luominen sekä suorituskyvyn jatkuva parantaminen ovat tärkeitä askeleita kohti huomispäivän tuottavampaa ja sitä kautta kestävämpää julkishallintoa. Datalähtöisen kulttuurin voivat saada toimimaan ainoastaan julkishallinnossa työskentelevät ja työnsä merkityksellisyydestä motivoituneet ihmiset, jotka sekä ymmärtävät oman työnsä perimmäisen palvelutarkoituksen ja jotka ovat valmiita datalähtöisyyden edellyttämään asennemuutokseen.

Ihmisten kouluttaminen, motivoiminen ja vastuuttaminen tähän uuteen organisaatiokulttuuriin on suuri haaste ja muutos, jota on lähestyttävä kokonaisvaltaisesti pelkkien inkrementaalisten organisaatiomuutosten sijaan. Datalähtöisen muutoksen käynnistäjiksi ja suunnannäyttäjiksi tarvitaankin kokeneita ja vastuutettuja tiedolla johtamisen pioneereja, joilla on vahvan datalähtöisen näkemyksen lisäksi kyky motivoida ihmisiä ja organisaatioita pitkäjänteisesti.

4.2 Kansalaisille yksilöllisempää julkista palvelua

Eräs julkishallinnon kestopäätöksiä on tarjota kustannus- ja palvelutehokkaampia palveluita kansalaisilleen eli säästää nykyisten palveluiden kustannuksissa ja/tai parantaa palveluiden vaikuttavuutta kustannusrakenne säilyttäen. Keskeinen keino parantaa palvelun vaikuttavuutta ja kansalaistyytyväisyyttä on segmentoida kansalaiset Big Datan tarjoamien menetelmin ja siten tarjota heille henkilökohtaisempaa palvelua.

Datalähtöisemmät palvelut vievät myös kansalaisten kokeman loppukäyttäjäkokemuksen seuraavalle tasolle. Sen perusteella millainen kansalainen on, missä kansalainen on, mitä

kommunikointivälinettä hän käyttää ja mitä hän on parhaillaan tekemässä, voi älykäs palvelu ehdottaa kansalaiselle ”päätöksiä” siitä, mitä hänen mahdollisesti tai luultavasti kannattaisi tehdä seuraavaksi. Kansalaisen tehtäväksi jää sitten tyypillisesti ja minimissään annettujen datalähtöisten ehdotusten hyväksyminen, tarvittaessa myös käyttäjälle päätöksen yhteydessä

annettujen tarkempien perustelujen tai toimentasuositusten pohjalta. Henkilökohtaisemmat ja päätöksentekoa helpottavat käytettävät palvelut parantavat automaattisesti kansalaisten tyytyväisyyttä ja luottamusta julkishallintoon.

Saksassa työvoimahallinto löytää big datalla töitä

Esimerkkinä suuren mittakaavan kansalaisten analyttisestä segmentoinnista ja julkishallinnon personoidusta asiakaspalvelusta on Saksan valtiollinen työllisyysvirasto Bundesagentur für Arbeit (BA), jonka päätehtävänä on integroida työnhakijat mahdollisimman nopeasti takaisin työvoimaan ja omalta osaltaan tehostaa yritysten rekrytointiprosesseja.

Segmentoimalla tarkasti tarjolla olevat työtehtävät ja työntekijät sekä räätälöimällä yksilöllisesti mm. työnhaun edistämiseen tarkoitetut neuvontapalvelut, BA on saanut aikaan merkittäviä tuloksia. Vuodesta 2006 vuoteen 2011 BA:n asiakkaan keskimääräinen työttämänäoloaika putosi 162 päivästä 136 päivään, työttömien määrä putosi 4,5 miljoonasta 2,9 miljoonaan ja BA:n kautta työllistyneiden määrä (per vuosi) kasvoi 240 000 ihmisestä 510 000 ihmiseen. Myös BA:n asiakastyytyväisyys niin työntekijöiden kuin työnantajien puolella parantui huomattavasti kyseisenä aikana.

(McKinsey Interview 2013, Frank Jurgen Weise, Behind the German jobs miracle.

http://www.mckinsey.com/features/government_designed_for_new_times/table_of_contents)

Laadukkaiden palveluiden tulee lisäksi kaikin tavoin säästää kansalaisten aikaa. Esimerkiksi kansalaisille näkyviä hakemus- ja tiedonsyöttöprosesseja tulisi automatisoida byrokratian ja käsittelyvirheiden vähentämiseksi. Kansalaisille on oltava mahdollisuus itse päivittää esim. kaikkia julkiseen asiointiinsa liittyviä, valmiiksi esitetyttä henkilö- ja hakemustietojaan. Riippumatta mikä taho julkiset digitaaliset palvelut varsinaisesti teknisesti rakentaa, tulee kaikkien palveluiden tukea helposti saatavissa olevaa reaaliaikaista palautetta niiden käytöstä, niin kansalaisten antaman yhteisöllisen palautteen kautta kuin palvelun varsinaiseen käyttöön perustuvaan analytiikan muodossa. Käytöstä kerättyä palautetta voidaan jälkeen päin analysoida niin palvelun käytettävyyden kuin varsinaisen julkisen asioinnin kehittämisen kannalta.

Julkishallinnon mahdollisia Big Data-käyttötapauksia

Julkishallinnolla on suuria määriä dataa, tyypillisesti rakenteisessa teksti- ja numeromuodossa. Julkishallinnon tuottavuutta voidaan parantaa yksittäisillä ja kustannussäästöjä teknisillä Big Data-ratkaisuilla, jotka on kohdistettavissa julkishallinnon eri hallinnonalojen ja yksiköiden täsmällisiin tietojenkäsittelyllisiin ongelmiin. Hyödynnettäessä Big Dataa julkishallinnossa, on teknisiä IT-ratkaisuja huomattavasti tärkeämpää saada aikaan datalähtöisyyden edellyttämä asennemuutos julkishallinnon ihmisissä ja työskentelytavoissa ja sitä kautta luoda jatkuvaan mittaamiseen ja hallinnon parantamiseen liittyvä datalähtöinen organisaatiokulttuuri. Alla on lueteltu mahdollisia sovelluskohteita Big Data -ratkaisuille:

Työvoima ja kouluttautuminen

- Koulutustarjonta-, työtarjonta- ja työvoimatarjontatietovarantojen yhdistäminen ja analyttinen segmentointi työttömän työvoiman mahdollisimman nopeaksi
- ja tarkaksi kohdistamiseksi tarjolla oleviin ja työttömille hakijoille yksilöllisesti soveltuviin työ- ja jatkokouluttautumistilaisuuksiin. Tällä voidaan
- omalta osaltaan tukea hallituksen tavoitteita nuorisotyöttömyyden torjumisessa, työurien pidentämisessä, ammatillisen koulutuksen tarjoamisessa,
- välityömarkkinoissa sekä työtarjousten lisäämisessä (ja niiden tiukennetusta vastaanottovelvollisuudesta).
- Sosiaalietuuksien ja -tukien väärinkäytön data-analyttinen ennakoiminen ja tukihakemusten/päätösten käsittelyn nopeuttaminen
- Tarjotun työn ja sen vastaanottamisen tarpeen perusteleminen data-analyttisesti tarjotusta työstä kieltäytymisen vähentämiseksi
- Education to Employment (E2E) parhaiden käytäntöjen soveltaminen yhdessä yritysten kanssa, mm. kasvan teknologiatyöttömyyden suunnan ennakoimiseksi
- Yleisesti työllisyyteen liittyvien koulutus- ja osaamistarpeiden tilannekuvaseuranta ja työvoimapolitiittisten strategioiden paremmaksi ennakoimiseksi

Harmaan talouden ja kansainvälisen veronkierron torjunta

- Tehostettu petosten ja virheiden havainnointi Big Data-ratkaisuilla osaksi harmaan talouden torjuntaohjelmaa
- Kovaan dataan perustuvien veroanalyysien täydentäminen esim. sosiaalisen median pehmeämmällä datalla

Hallituksen kestävän kehityksen mittarit

- Yhteiskunnalliseen kestävyys liittyvät Big Data-mittarit ja niiden kansalaisystävällinen seuranta
- Ympäristön kestävyys liittyvät Big Data-mittarit ja niiden kansalaisystävällinen seuranta

Kansalaisille avoimempi ja ymmärrettävämpi kuva valtiontalouden tilasta

- Kansantalouden data-analyttinen "simulaattori" palvelemaan yksityisten, julkisten tahojen ja yksittäisten kansalaisten tekemiä tutkimuksia koskien
- valtiontalouden tilaa
- Valtion rahoitusvirtojen ihmisystävällisemmät visualisoinnit avoimen datan pohjalta, esim. oman äänestyspäätöksen helpottamiseksi
- Muut taloudelliseen kestävyys liittyvät Big Data-mittarit ja niiden kansalaisystävällinen ja -motiveiva seuranta

5. Läpileikkaavia teemoja

Läpileikkaavat teemat ovat asioita, jotka vaikuttavat useilla sovellusalueilla. Teemoilla on oma tärkeä osansa big data-kehityksessä. Teemat ovat vielä pitkälti kehittyviä, ripeät liikkeet ja kehitys niissä voivat tarjota Suomelle monia mahdollisuuksia.

5.1 Mydata

My Data viittaa toimintakulttuuriin, jossa organisaatio palauttaa keräämäänsä yksilökohtaista tietoa ihmiselle itselleen. Sen jälkeen ihminen voi hyödyntää sitä suoraan tai jakaa sen edelleen haluamallaan tavalla. Edelleen jaettu tieto voi olla analysoimatonta raakadataa tai se voidaan jalostaa esimerkiksi visualisoimalla muotoon, jossa muut voivat sitä hyödyntää.

My Data -näkökulmat ovat tärkeitä big dataan liittyvissä keskusteluissa, koska niiden avulla voi purkaa big dataan ja yksityisyyteen liittyviä uhkia. My Data -aloitteiden päämääränä on ollut henkilöitä koskevan tiedon ympärille syntyvien sovellusten, palveluiden ja toimintarakenteiden kehitys tavalla, jossa ihmisillä on keskeinen päätösvalta tiedon keräämisen, jalostamisen ja hyödyntämisen suhteen. My Data -aloitteesiin liittyvä keskustelu yksityisyydestä korostaa ihmisten valtaa päättää siitä, kuka heidän tietojansa saa käyttää ja uusiokäyttää sen sijaan, että ihmiset antavat kerralla suostumuksen aineistojen käyttöön. Digitaalisuuteen liittyy nimenomaan mahdollisuus monistaa ja käyttää tietoa yhä uudelleen, jolloin ihminen ei suostumustaan antaessaan voi tietää tai hallita tulevaisuuden käyttötarkoituksia.

Onnistunut My Data -työskentely edellyttää, että yhteisistä henkilökohtaisiin aineistoihin liittyvistä periaatteista pystytään sopimaan. Tarvitaan viisasta sääntelyä, jonka avulla aineistojen jakaminen on mahdollista. Tarvitaan koneluettavia rajapintoja aineistojen jakamisen tueksi, standardeja ja palveluita tiedon hallittuun siirtämiseen, varastointiin, käsittelyyn ja analysointiin. Nämä pyrkimykset tukevat myös big datan hyötykäyttöä tulevaisuudessa.

//KUVA BIG DATA KEHIKKO//

5.2 Quantified self

Quantified Self -liikkeen juuret ovat Kaliforniassa ja erityisesti Wired-lehden ympärillä. Liikkeen keskeisenä ajatuksena on yksilöllisen mittaustiedon hyötykäyttö terveyden ja hyvinvoinnin edistämiseksi. Viime vuosien aikana liike on levinnyt eri puolille maailmaa, ja saanut uusia muotoja. Kyse ei ole enää yksinomaan superterveydestä haaveilevien teknointoilijoiden liike vaan mittaamista istutetaan yhä uusille elämäalueille. Big data ja QS -näkökulmilla on paljon yhteistä, koska QS on yksi mahdollinen big datan tuottaja. Digitaalisten itsenmittausaineistojen arkistointiin tarvitaan uusia ratkaisuja. Tiedon hallinta on turvattu, kun yksilö omistaa tietonsa ja antaa sille käyttöoikeuden

halutessaan. Kun itsenmittaajat haluavat jakaa aineistoja myös muille, tiedon käyttö on turvattava tavalla, joka hyödyttää sekä tiedon jakajaa että hyödyntäjää.

Itsenmittauksen ympärillä on paljon uusia aloitteita, joita motivoi myös terveydenhuollon toimialamuutos. Suomessa itsemittautusteknologioita kehittäneitä, kansainvälisillä markkinoilla toimivia yrityksiä on useita, muun muassa Firstbeat, Suunto ja Polar. Lisäksi alalla on lukuisia lupaavia start up -yrityksiä. Mittalaitteiden avulla ihmiset saavat uudenlaista tietoa itsestään. Itsehoito on aiempaa helpompaa, kun ihmiset voivat mitata kotona verenpainettaan tai seurata unen laatuaan. Tämä voi muuttaa vähitellen terveydenhuollon painopistettä ja lääkäreiden työtä. Lääkärit voivat keskittyä tiukemmin nimenomaan erikoistuneeseen sairauksien hoitoon, kun sairauksien ennaltaehkäisyyn on uudenlaisia vaihtoehtoja. QS-ajatus on leviämässä myös hyötykäyttöön ympäristön tilan arvioinnissa ja hoidossa.

Toisaalta QS haastaa myös käsityksiä terveydenhallinnasta ylittämällä olemassa olevan terveystoimialan. Itsenmittaajat ovat osoittaneet, että esimerkiksi ajankäyttödata voi olla terveysdataa. Esimerkiksi riippuvuuksia voi tarkastella ajankäytön avulla. Kiinnostavaa on myös terveydentila suhteessa ympäristöntilaan. Voiko ilmanlaatuista ennakoita tulevaisuudessa keuhkohtaumapotilaan kohtauksen?

QS-näkökulma on alkanut levitä myös muille toimialoille, esimerkiksi oman talouden hallintaan tai koulutukseen. Mittatietoa voi käyttää esimerkiksi omaehtoisen oppimisen seurantaan tai tekemään näkyväksi opettajien pärjäämistä eri luokkien kanssa.

5.3 Joukkoistaminen

Demokraattista osallitumisen mahdollisuuksia sekä kansalaisten äänen kuulumista esimerkiksi päätöksentekoon voidaan lisätä big datan avulla. Teknologia ja uudet tiedon analysointimenetelmät tulisi valjastaa tukemaan myös demokratian ja kansalaislähtöisyyden kehitystä. Tämä olisi esimerkillistä ja rohkeaa, toisaalta kokemukset voisivat olla myös vientikelpoisia. Ylipäänsä asiakkaiden, käyttäjien sekä kansalaisten osallistaminen on tarpeellinen lähtökohta big data-kehitykselle. Välttämättä ihmisten tietoja ei vain viedä, kuten synkemmissä big datan riskejä korostavissa

Kansalaismielipide päätöksenteon tueksi big datan avulla?

Big data tarjoaa paljon mahdollisuuksia myös osallistamisen ja demokratian kannalta. Uusien menetelmien avulla voitaisiin esimerkiksi käsitellä erilaisia kansalaisten mielipiteitä tarjoavia aineistoja ja lähteitä.

Muun muassa sosiaalisen median, erilaisten kyselyiden ja keskusteluiden yhdistelyllä voisi saada paremman kuvan kansalaisten mahdollisista ongelmista ja esimerkiksi reaktioista erilaisiin päätöksiin. Menetelmien avulla voisi löytää piileviä aiheita ja tehdä päätöksenteosta ja hallinnosta responsiivisempää sekä kansalaislähtöisempää.

arvioissa mainitaan, vaan ihmisiä voidaan big datan avulla myös osallistaa mukaan esimerkiksi heitä koskevien päätösten tekemiseen. Tässä lähtökohtana on käyttäjän suostumus.

Samalla joukkoistaminen voi olla toimiva menetelmä myös uusien datalähtöisten palveluiden kehittämisessä. Käyttäjien havaintoja tai heiltä saatua

tietoa voi hyödyntää toiminnan parantamisessa ja osittain erilaisten mallien synnyttämisessä. Esimerkkinä tästä voisi olla vaikka reittisuosituksia yms. antava liikennesovellus, joka yhdistelisi sekä käyttäjien dataa että esimerkiksi reaaliaikaista liikennedatata saavuttaakseen mahdollisimman tarkat ennusteet sekä paremman palvelun.

Osallistamista sekä käyttäjien havaintojen keräämistä ja hyödyntämistä tulisi soveltaa myös hallinnon palautteen keräämisessä ja toiminnan suunnittelemisessa yhä enemmän.

5.4 Tiedon etsintä

Tietomäärien valtava lisääntyminen luo päivittäisiin työtehtäviin lisähaasteita myös kaikille tietotyöntekijöille. Esimerkiksi tiedonetsintä on alue, jota paremmin tukemalla voidaan saavuttaa merkittäviä kustannussäästöjä. Erilaisille tietotulvaa helpottaville työkaluille ja mm. etsintämentelmille on monkansallisten jättien hallitsevasta asemasta

SCINET tekee parempia hakuja

HIITissä on kehitetty SCINET-tiedonhakujärjestelmä, joka on testeissä todettu Google Scholaria tehokkaammaksi. SCINET hakee tällä hetkellä tieteellisiä artikkeleita. Konseptia ollaan kuitenkin laajentamassa myös muunlaiseen tiedonhakuun.

http://www.tiede.fi/artikkeli/uutiset/suomalaistutkijat_tekivat_uudenlaisen_hakukoneen.

<http://research.ics.aalto.fi/mi/online-papers/Ruotsalo13cikm.pdf>

huolimatta vielä kysyntää esimerkiksi tutkimuksen parissa. Tietotyö sinällään lisääntyy jatkuvasti, joten erilaisiin tarpeisiin räätälöidyllä, erilaisen tiedon etsintää helpottavia työkaluja tarvitaan monilla aloilla. Suomessa tehdään merkittävää alaan liittyvää tutkimusta.

Alalta voi oikeilla panostuksilla löytyä hyvin erilaisia hyötyjä. Tietyn alan työhön kuuluva tietotulvan

hallinta ei välttämättä onnistu vielä nykyisillä työkaluilla. Uusilla tiedonhakumenetelmillä, esimerkiksi vuorovaikutteisesti oppivan visuaalisen järjestelmän, avulla voidaan löytää relevantimpaa tietoa helpommin ja nopeammin. Näin työn laatu ja tehokkuus paranee.

5.5 Paikkatieto

Paikkatiedoilla tarkoitetaan tietoja, joilla selvitetään jonkin tietyn paikan ominaisuuksia tai joita käytetään tiettyjen paikkojen etsimiseen. Tähän asti paikkatietoja on hyödynnetty luonnonvarojen kartoittamisessa, maankäytön suunnittelussa ja ympäristönsuojelussa. Paikkatiedot kuitenkin tarjoavat lukemattomia sovellusmahdollisuuksia lähes kaikilla liiketoimintasektoreilla ja julkisen sektorin hankkeissa.

SRV ja paikkatiedot:

Rakennus- ja kiinteistöalan monialayhtiö SRV on käyttänyt paikkatietoja liiketoimintansa suunnitteluun, analysointiin ja optimointiin päätöksenteon tukena jo useita vuosia. Yrityksen tarpeet ja ajankohtainen, jatkuvasti päivittyvä tieto saivat SRV:n kehittämään toimintatapaansa edelleen. Tarkoituksena oli saada yhdistettyä jatkuvasti päivittyvät tietosisällöt omiin aineistoihin, jotta näitä voitiin analysoida yhdessä. Yritys päätyi valitsemaan pilvipalvelunjolla, jonka avulla esimerkiksi Tilastokeskuksen ja maanmittauslaitoksen taustakarttoja voitiin käyttää laajasti hyödyksi.

Avoimeen dataan perustuvat karttapalvelut antoivat laajat mahdollisuudet yhdistää jo olemassa olevat tiedot ajanmukaisiin karttoihin ja esittää saadut tulokset dynaamisen selainsovelluksen avulla graafisesti.

http://www.esri.fi/referenssit/srv-helppokayttainen_paikkatietojarjestelma_kiinteistoliiketoiminnan_tukena/

Paikkatietojen innovatiivinen käyttäminen tarjoaa aivan uusia näköaloja kuljetuspalveluiden, yhdyskuntasuunnittelun ja liiketoiminnan suunnittelun suhteen. Palveluverkostojen optimaalinen kohtaavuus palveluja haluavan kanssa voidaan suunnitella täysin vedenpitävästi. Yksinkertaisimmillaan paikkatiedoilla voidaan kohdentaa palveluista kiinnostuneelle mahdolliselle käyttäjälle tai asiakkaalle täsmällistä informaatiota tarjolla olevista lähipalveluista.

Paikkatiedoilla voidaan kiinnittää huomiota mm.

ihmisten kulutustottumusten esiintymiseen kartalla ja luoda näiden tietojen perusteella malli joukon ominaisuuksista. Palveluyritys voi paikkatietojen perusteella määritellä optimaalisen sijainnin yritykselleen, jonka toimintasäteen tai tarjotun palvelun sijaintialueella asuu optimaalinen määrä palvelun kohderyhmään kuuluvia ihmisiä. Sama metodi sopii myös tiestön suunnitteluun sijainnin optimoimisen avuksi ja tulevaisuuden ennakkoinnin apuvälineeksi.

6. Toimenpiteet

6.1 Datatietoisuus

Big datan hyödyntäminen kaikilla sektoreilla on olennaista, joten tietoisuutta aiheesta on lisättävä kohti datan tunnistamista, kokeiluja ja kehitystoimintaa. Sanotaan, että big datan teknologiakehitys on evoluutiota, mutta samalla liiketoimintaprosessien kehitys käy läpi nyt revoluutiota. Erityisesti yrityksissä tarvitaan siten toimenpiteitä ymmärryksen lisäämiseksi liiketoiminnan muutoksesta dataintensiivisessä kilpailussa.

Suomen jokaisen organisaation päätöksenteon tulisi muuttua tietoon perustuvaksi (data driven) sen sijaan että päätökset tehtäisiin mutu-pohjaisesti. Maailmalla on tietopohjaisesta päätöksenteosta esimerkkinä mm. amerikkalainen autonvalmistaja Ford, joka selvisi autoteollisuuden kriisistä siirtymällä data-driven päätöksentekoon. Tällainen

tietoon perustuva päätöksenteko vaatii tuekseen toimivan tiedonkeruun ja analysoinnin. Sama tiedonkeruu ja -analysointi palvelee myös open data -aloitteita.

- a) **”Datalähettiläitä”** hankitaan tuomaan big data- tietoutta ja osaamista yritysten ja hallinnon käyttöön toteutettuna mm. kansainvälisinä vierailijaluentoina ja osajavaihtona
- b) Luodaan alan yrityksiä, tutkijoita ja asiantuntijoita kokoava avoin Big data-klusteri sekä osaamisen kehittämistä tukevaa verkostotoimintaa tämän ympärille (Tekes, Teknologiateollisuus, Tekniikan Akateemiset, Ohjelmistoyrittäjät, TTL, yritykset jne).
- c) Julkisen hallinnon organisaatioihin nimetään datavastaavia, joiden tehtäviin kuuluu datan keruun ja analysoinnin järjestäminen. Yksityisiä yrityksiä ja muita yhteisöjä varten voidaan järjestää neuvontaa.

6.2 Koulutus

Osaamisen kehittäminen on olennaisin big dataa edistävä ja samalla kehitystä rajoittava tekijä. Tarvitaan sekä teknistä että toimialaspesifistä osaamisen kehittämistä. Osaamisessa tärkeää erilaisten datamuotojen käsittely ja jalostaminen (”blended human and algorithmic data processing”). Datatieteilijältä odotettavat taidot (kuvailua edellytyksissä) ovat poikkitieteellisiä, joten tiimien (osaajapoolit) rakentaminen voisi olla yhtenä ratkaisuna osaajapulaan.

Tutkintoperusteisen ja käytännönläheisen opiskelun kehittäminen esimerkiksi harjoitteluna yrityksissä, osallistumisena tutkimusohjelmiin tai vapaaehtoisina kursseina tai verkkomoduleina ovat molemmat tarpeen.

Data-analytiikkaosaamista on saatava kaikille oppitasoille peruskoulusta lähtien. Yliopistotutkinnon yleisiin/yleissivistäviin opintoihin tulisi saada datan hallinnan ja käsittelyn painotusta oppiaineesta riippumatta.

Alan koulutusorganisaatiot, yritykset ja asiantuntijat muodostavat opetushallituksen johdolla näkemyksen tarjolla olevasta koulutuksesta ja tarvittavista vähimmäistaidoista eri rooleissa, jotka tulevat mukaan tutkintoihin (ammattikorkea, yliopisto, täydennyskoulutus). Oppilaitokset ja yliopistot voivat täten myös paremmin erikoistua ja tehdä yhteistyötä esimerkiksi verkkomuotoisen opiskelun kehittämiseksi.

EU:n tavoitteena on luoda Euroopassa yliopistojen verkosto datataitojen osalta.

- d) **Vaikutetaan siihen, että peruskoulun opetussuunnitelmiin sisältyy dataan käyttöön liittyvää opetusta vuonna 2016 ja tuetaan open source teknologioiden leviämistä.**

Opettajankoulutuksen kehittäminen (big data opetustaidot):

- e) Kartoitetaan opetushallinnon johdolla tarvittavia osaamisprofiileja ja olemassa olevaa koulutustarjontaa.**
- f) Koordinoidaan kansallisesti käynnisteillä ja suunnitteilla olevia big data koulutusohjelmia.**
- g) Osallistutaan EU-laajuisen big dataan erikoistuvien yliopistojen verkoston luomiseen**

6.3 Soveltava osaaminen

Soveltavan ja täydennyskoulutuksen malleja on kehitettävä big data asiantuntijaksi kasvamiseksi. Yrityksissä osaamista kehittäisivät erilaiset harjoittelujaksot ja projektit yms. alan opiskelijoille, tutkijoille ja työmarkkinoilla vapaana oleville osaajille.

- h) Ohjelmointi/tietojenkäsittely/bisnesprosessien opiskelijoita kannustetaan yrityksiin big data harjoitteluun, projekteihin ja opinnäytetöihin (hyvänä esimerkkinä Aallossa diplomityöparit)**
- i) Toteutetaan reaaliaikaista tarvekartoitusta siitä, kuinka paljon ja millaista osaamista yritykset tarvitsevat (big data yritysklusteri/foorumi)**
- j) Täydennys- ja muuntokoulutusta uudistetaan vastaamaan big data osaamistarpeita**
- k) Tuetaan yritysten big data osaamista kehittävien sisäisten hautomojen toimintaa, osaamisvaihtoa ja työnkiertoa**

6.4 Tutkimus

6.5 Infrastrukturi

Big datan hyödyntäminen vaatii korkealuokkaista tieto- ja viestintäinfrastruktuuria. Varmistetaan, että perusedellytykset, kuten viestintä- ja tiedonsiirtoyhteydet, pilvikapasiteetti yms. ovat kansainvälistä huipputasoa ja ennakoidaan tarpeita infrastuktuurin osalta niin yksittäisen toimijan kuin yhteishankkeiden/toiminnan kannalta.

Keskeisten kansallisten toimijoiden tunnistaminen ja organisointimallin luominen on tarpeen myös infrastruktuurin kehittämistä ohjaamaan. Infrastruktuurikoordinaation tehtävä on hakea tehokkuutta ja yhteentoimivuutta.

Kansainvälisesti hyvänä esimerkkinä voidaan verrata Open Cloud Massachusettsin toimintatapaa ja datapalvelujen "biotope"-ajattelua sekä UK:n ODI toimintaa muun muassa datasovellusten edistäjänä ja datan saatavuuden kehittäjänä.

Jaettujen infrastruktuurien kehittämistä tehdään tarpeiden mukaan ja ne voivat tarkoittaa esimerkiksi: - Tiedonkäsittely- ja laskentainfrastruktuurien kehittämistä - Yhteisiä datan hallintasysteemejä ylläpitoon ja laadunvarmistukseen - T&k-yhteistyön ja arvioinnin kehittämistä erilaisista analyysimenetelmistä - Datapalvelujen ekosysteemin luontia, kuten datan jakelupisteet, kauppapaikat ja julkaisukirjastot sekä datasovellusten edistäjärooli - Kokeilulaboratorio, jossa voidaan testata teknologioita ilman omaa mittavaa panostusta

l) Kehitetään kansallista big data infrastuktuuria verkostomaisella yhteistyöllä

6.6 Tekniset käytännöt ja standardit

Big datan teknologisia kehitystarpeita on tunnistettu olevan mm. datan varastoinnin tekniset standardit, datan välittämisen ja integroinnin tavat. Teknologiakohtaisia ekosysteemejä on syntymässä, joten näiden ymmärtäminen ja sovittaminen omaan kehitysympäristöön on olennaista. Datan käsittelyn ja hallinnoinnin harmonisointi edistäisi datan yhteen toimivuutta ja siten käyttöä. On huomattava, että dataan liittyviä standardeja (esim INSPIRE, PSI) on jo olemassa, joten näitä voidaan hyödyntää ja soveltaa myös big dataan.

m) Tunnistetaan keskeiset standardointiprosessit joissa tulee olla mukana sekä organisoidutaan toimimaan näissä

n) Standardien luominen erityisesti datan varastointiin (storage) ja siirtoon (exchange) kehittää datan käsittelyn tekniikoita, työskentelymetodeja ja tehokkaita algoritmeja.

o) Tietosuoja ja yksityisyys on myös teknologinen kehityskysymys. Panostetaan tieto/yksityisyydensuojan ratkaisujen teknologiakehitykseen (esim. Tekes)

6.7 Datan käytettävyys ja avoimuus

Avoimet rajapinnat ja data ovat myös big datan hyödyntämisen mahdollistajia, joten kehitystyötä tulisi edelleen vahvistaa sekä luoda tapoja yhdistää julkista avointa dataa ja yritysten dataa

p) Datan hyödyntämiseen ratkaisuja etsiviä ja kehittäviä henkilöitä rekrytoidaan hallintoon, esimerkiksi kumppanikoodarimallin mukaisesti, College-to-govt harjoitteluohjelmat tms.

- q) Yhteentoimivuuden kehittäminen datan jakamisen standardeilla ja yhteishankkeilla erityisesti kunta- ja kaupunkiympäristössä (mm Kuusaika- hankkeen puitteissa)**
- r) Määritellään kansallisen dataportaalin hyödyntäminen myös big data kanavana Avoimen tiedon ohjelman osana (VM)**
- s) Velvoitetaan virastoja tunnistamaan data-aineistoja, joista big data kehityksen kannalta saataisiin merkitäviä hyötyjä. Tuetaan tässä virastojen pilottiprojekteja (Avoimen tiedon ohjelma/kehukset).**

6.8 Sääntely

Suurena haasteena on se, miten tasapainottaa big datan hyödyntämiseen ja tietosuojaan liittyvät toimet sääntelyssä. Henkilöiden yksityisyyden suoja ei saa tässä vaarantua.

Sääntelyn kehittämisen ajatuksena tulee kuitenkin olla se, että markkinoille tulon esteitä voidaan purkaa ja samalla luodaan kannusteita datan saatavuuteen huomioiden henkilöiden oikeudet dataan. Yleisesti säädösvalmistelussa tulisi huomioida dataan liittyvät edellytykset, joten voidaan myös kysyä onko big datassa erityispiirteitä tai tarpeita huomioitavaksi sääntelyssä.

Sääntelyä tukisi yhteiset käytännöt ja periaatteet, ns. Big data "etiketti" siitä, miten dataa hyödynnetään kunnioittaen niin kansalaisten kuin yritystenkin oikeuksia. Hyviä datan hallinnon ja käsittelyn tapoja tulee luoda yhteistyössä. Tämä voisi toimia myös Suomen vahvuutena kansainvälisessä kilpailussa datavarannoista.

- t) Tehdään selvitys big datan käyttöön vaikuttavista laeista**
- u) Varmistetaan, että EU:n tietosuoja-asetus ja sen soveltaminen Suomessa mahdollistaa big data kehityksen yksityisyydensuojaa vaarantamatta**
- v) Kehitetään suomalainen Big data "etiketti" ohjesääntö**

6.9 Yhteistyö ja datan vaihto

Yleisesti jaetaan näkemys tarpeesta hakea datan avulla synergioita eri aloilta. Pilvipalvelujen hyödyntäminen on tässä olennaista, sillä se mahdollistaa uudenlaisen datan yhdistelyn ja jakamisen. Löydettävä siten toimia, joilla erityisesti yritysten horisontaalista ja vertikaalista verkottumista voidaan edesauttaa siten, että (big ja avointa) dataa jaettaisiin sopivalla alustalla/rajapinnoilla yhteiseen käyttöön ja yhdisteltäväksi uudella tavalla.

Yrityksillä (erityisesti suurilla) on hyödyntämätöntä dataa, mutta ei osaamista/resursseja/ideaa käyttää tätä. Vaatii toimien suuntaaminen: 1. kannustetaan/tuetaan yrityksiä avaamaan rajapintoja dataansa 2. yhteiset kehitysalustat datan jakamiseksi, joista jalostuu uusia lisäarvopalveluja

Haasteena nähdään se, ettei yrityksillä ole välttämättä kannusteita tai liiketoiminnan intressien kannalta järkevää tapaa jakaa dataansa. Kuitenkin "varallisuus"-ajattelun riskinä voi olla, että käyttämättömästä datasta ei saada mitään arvoa ja kilpailijat menevät ohi. Yhteistyön kehittäminen vaatii uutta ymmärrystä jakamistalouteen ja siten hyödyn- ja tulonjakomallien kehittämistä. Datan markkinapaikkoja tulee myös kehittää.

Big datan potentiaalista kertoo jo se, että perusteilla on big data labseja ja tutkimuskeskuksia ympäri Eurooppaa, kuten esimerkki Saksan Smart Data Innovation Labista kiellii. Suomen on mentävä mukaan tähän toimintaan, jotta yhteistyö, kokeilut, tutkimus, yritysten sovellusosaaminen ja siten yleinen alan kehitys vauhdittuisi. Osaavien ihmisten välisten yhteyksien rakentaminen katsotaan nyt olennaiseksi, samoin kuin hyvien käytäntöjen jakaminen eri toimijoiden välillä. Verkostotoiminnan tavoitteiden on oltava kuitenkin selkeitä: toiminnan on tuettava yritysten ja ihmisten suorituskyykyä tuloksia seurattavissa olevalla tavalla.

w) Selvitetään tarvetta Big data-hubille, erityisesti isojen ja pk-yritysten yhteisille datahankkeille ja datan vaihdolle liiketoiminnassa. (DIGILE, Teknologiateollisuus, yritykset)

Singapore ja social analytics: social analytics-yhteistyökonsortio kehittää alan työkaluja yhdessä.

<http://www.ida.gov.sg/Collaboration-and-Initiatives/Collaboration-Opportunities/Store/Social-Analytics-SA-for-Business-Enterprises-Call-for-Collaboration-CFC>

- x) Edistetään erityisesti pk-yritysten ja start-uppien big data resursseja (välineitä ja pääsyä aineistoihin) ja osaamista. Aktivoidaan näitä yrityksiä myös mukaan kansallisiin hankkeisiin ja yritysten aloitteisiin (vrt. GE:n tarjoamat fasilitetit)**
- y) Luodaan kansallinen data-analytiikan toimijoiden verkosto (tutkimuslaitos/yliopistoyhteistyö), jossa edistetään myös teknologiasiirtoa yrityksiin (menetelmät, työkalut, osaaminen)**
- z) Osallistutaan ja jaetaan kansallisesti näkemyksiä EU:n Big/NESSI data foorumin toimiin**

EU:n big data public private forum

Olemassa olevaan infrastruktuuriin perustuva ekosysteemi, jossa luodaan myös yhteistyön malleja, fasilitoidaan datan siirtoa (data exchange) ja tarkastellaan samalla miten liiketoiminta/(asiakas) johtaminen muuttuu. Forumilla tehdään kokeiluja big data aineistoilla. Niiden avulla kehitetään osaamista, tutkimuksen hyödyntämistä ja työkalujen siirtoa (research transfer community) <http://www.big-project.eu/>

Verkottuneet käyttäjät (connected users) ovat nyt big data kehityksen keskiössä, sillä se tarkoittaa useilla sektoreilla tiedon hyödyntämisen nopeutumista. Esim. liikenteessä käyttäjiltä kerättyä tietoa yhdistetään liikenteen hallintaan.

- ä) Kehitetään joukkoistamiseen perustuvia big data- malleja, joista parhaillaan syntyy uudentyyppisiä laajoja ekosysteemejä.**

6.10 Kokeilut ja t&K rahoituksen suuntaaminen

Kokeilevaa toimintaa big datan hyödyntämiseksi ja analyysimenetelmien soveltamiseksi käytännön tilanteisiin ja yhteiskunnallisiin haasteisiin sekä uuteen liiketoimintaan tulee tukea ja edistää. Big datan kokeilut tulisi saada siten vahvemmin mukaan erilaisten ohjelmien ja kehityshankkeiden osaksi.

Esimerkiksi Tekes on käynnistänyt ohjelmavalmistelun Teollisen Internetin alueelle. Saman valmistelun yhteydessä tarkastellaan tarpeet myös 5G-mobiiliverkkojen alueella, joka on yksi tekijä teollisen internetin toteuttamisessa. Valmistelun tuloksena voi olla yksi tai useampi ohjelma tai jokin muu kokonaisuus. Big data ja data-analytiikka ovat myös keskeisessä roolissa teollisen internetin sovelluksissa. Valmistelun tuloksena mahdollisesti käynnistyvä(t) ohjelma(t) tai muut toimenpiteet voisivat käynnistyä syksyllä 2014 (Q4/2014). Valmistelun verkkosivut <http://www.tekes.fi/teollinen-internet-ja-5G>.

Eriyisesti menetelmiä ja teknologiaa kehittävät yritykset voivat tarvita pääomaehtoista rahoitusta kehitystoimintaansa. Pääomaehtoista rahoitusta ja big datan liiketoimintamahdollisuuksien arviointiosaamista tulee siten kehittää. Yhdysvalloissa big data houkuttaa merkittäviä riskisijoituksia. Suomalaisen menetelmä- ja teknologiaosaamisen kehittäminen ja kaupallistaminen tulee löytää niin yksityistä kuin julkista rahoitusta sekä näiden yhdistelmiä esimerkkinä Tekes Pääomasijoitus Oy.

- ä) Kansallista tukea ja verkostoa EU-tukiohjelmien hyödyntämiseen vahvistetaan erityisesti big data painotuksella (Horizon verkosto, Tekes, kehittäjäorganisaatiot)**

- ö) Tuetaan erilaisia big data kokeiluja ja jaetaan näistä saatuja kokemuksia. Eriyisen tarpeelliseksi katsotaan datan saatavuus hallituissa kokeiluissa, joihin liittyy esimerkiksi mobiilipaikantaminen**

- aa) Disruptio datan vaikutuksesta tulevat voimakkaasti myös liiketoimintamalleihin: kokeilevat kehityshankkeet hyödyn ja**

tulonjaonmallien kehittämiseksi (vrt. myös yritysten yhteistyön aktivointitoimet)

bb) Pääomaehtoisen rahoituksen (yksityiset ja julkiset pääomasijoittajat) saatavuutta tulee edistää

6.11 Mydata

cc) MyData selvityksen malleja testataan valituilla alueilla kehityksen vauhdittamiseksi ja uusien tietomallien yleistymiseksi eri alojen yritysten avulla. Pitkällä aikavälillä luodaan alustaa MyDatalle, johon yksilö voi omaa tietoaan kerätä ja haluamallaan tavalla jakaa sekä hyödyntää.

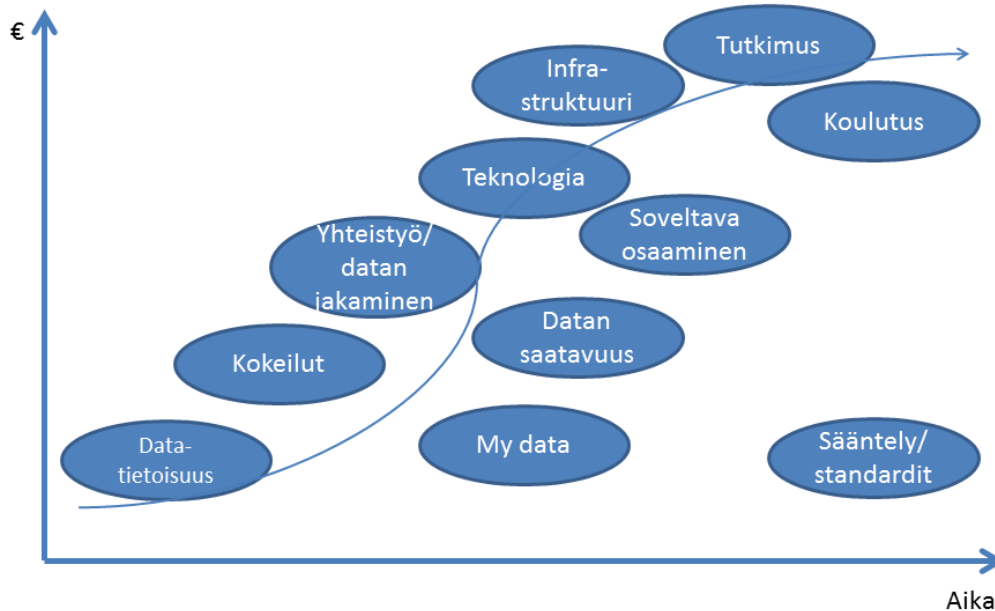
dd) Yhteistyö Iso-Britannian MiData-labin kanssa Midata-kehityksen vahvistamiseksi

7. Vaikuttavuus

Tässä osiossa on tarkoitus arvioida big datan potentiaalia kansantaloudessa ja ja tuoda esille näkökulmia siitä, miten strategiassa esitettyjä toimenpiteitä voitaisiin arvioida.

Alla olevassa kuvassa on alustavasti arvioitu strategiassa esitettyjen toimien vaikuttavuutta ajassa. Esimerkki kuvaa sitä, miten välittömästi toimet vaikuttavat: kokeilut vaikuttavat lyhyellä aikavälillä, koulutus varsin pitkällä ja niin edelleen. Toimeenpanossa on varmistettava, että saavutetaan vaikuttavuutta sekä lyhyellä että pidemmällä aikavälillä.

Toimien vaikuttavuus ajassa ja tarvittavat taloudelliset resurssit



Taustaksi tuodaan tähän arvioita big datan potentiaalista kansainvälisiä ja kotimaisia tutkimuksia tai selvityksiä vertaillen (OECD, WEF, EU). Suomen osuutta voidaan arvioida esimerkiksi bkt suhteella.

Strategiassa valitaan vaihtoehtoisia/toisiaan täydentäviä tarkastelukulmia vaikuttavuuteen toimenpiteiden tarkentuessa:

- Esitetään arvioinnin pohjalta, miten ja kuinka merkittävästi edellä mainittujen toimien toteuttaminen vaikuttaisi yhteiskuntaan tai talouteen. Linkitetään "grand challengeihin" mm. kestävyysvaje (esimerkkien kautta esim. julkinen hallinto)
- Määritetään arvio tarvittavista investoinneista
- Datan arvoketju ja arvonluonti A) Datan arvoketju: 1. data hankinta (acquisition) 2. analyysi 3. tuki ja joukkoistaminen (curation) 4. varastointi 5. käyttö B) Arvonluonti: miten datasta saadaan arvoa (data extraction)
- Kansantalouden tasolla tarvitaan esimerkkejä demonstroimaan big datan käytön vaikutuksia kasvuun ja työllisyyteen. Vaatii myös datan relevanttiuden arviointia: 1. datan määrällinen kysyntä (edustaa myös poikkisektorien mahdollisuuksia, kuten geodata) 2. kyseisen sektorin merkitys (osuus kansantaloudesta; esim teollisuus, palvelut, julkinen sektori, maatalous, kolmas sektori)
- Arvio Suomen big data PPP ekosysteemin edellytyksistä ja potentiaalista. Ensin on ymmärrettävä ketä tässä toimii ja mitä toimintaa on, joten olisi kartoitettava keskeiset toimijat ja hankkeet
- Potentiaali, että kehitystoimet toteutetaan versus vaihtoehto ettei asiaa edistetä mitenkään

- ee) **Tehdään taloudellinen ja yhteiskunnallinen analyysi big datan potentiaalista Suomessa ja strategisten toimenpiteiden vaikuttavuudesta.**

8. Strategian toimeenpano

Luodaan yhteistyössä tiekartta toimien vastuista, etenemisjärjestyksestä, aikatauluista, vaikuttavuudesta ja mahdollisista riskeistä. Vahvistetaan strategian toimeenpano valtioneuvoston periaatepäätöksellä ja/tai hallitusohjelmassa. Toimet, jotka voidaan käynnistää heti resursoidaan ja organisoidaan välittömästi.

- ff) **Strategian toimeenpanosta luodaan aikataulutettu ja vastuutettu suunnitelma.**

ⁱ <http://www.emc.com/leadership/digital-universe/index.htm?pid=home-dig-uni-090414>

ⁱⁱ <http://www.kapowsoftware.com/resources/infographics/intelligence-by-variety-where-to-find-and-access-big-data.php>

ⁱⁱⁱ http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

^{iv} http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf

^v <https://www.gov.uk/government/publications/uk-data-capability-strategy>

^{vi} Työryhmän työtä varten koottu katsaus muiden maiden big data-strategioista.
<http://www.lvm.fi/lvm-mahti-portlet/download?did=137476>

^{vii} ICT2015, <http://ict2015.fi/toimenpiteet/polku-10-big-data/>

^{viii} http://www.computerworld.com/s/article/9244363/How_analytics_helped_Ford_turn_its_fortunes

^{ix} <http://newsroom.cisco.com/press-release-content?articleId=888280>

^x Pepper, Garrity

http://www3.weforum.org/docs/WEF_GlobalInformationTechnology_Report_2014.pdf

^{xi} <http://www.technologyreview.com/news/527376/silicon-valley-to-get-a-cellular-network-just-for-things/>

^{xii} <http://www.theguardian.com/public-leaders-network/2014/apr/15/big-data-open-data-transform-government>

^{xiii}

http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf

^{xiv} Sydänmaanlakka, P. 2009. Jatkuva uudistuminen – luovuuden ja innovatiivisuuden johtaminen. Talentum. Hämeenlinna. s. 13.

^{xv} Nonaka, I., Toyama, R., Hirata, T. 2008. Managing Flow – a Process Theory of the Knowledge-Based Firm. Palgrave Macmillan. New York. 45.

^{xvi} Sydänmaanlakka 2009, 208.

^{xvii} Wikinomics-ajattelusta ks. Tapscott, Don & Williams, Anthony D. Wikinomics. 2006. How Mass Collaboration Changes Everything. Portfolio, New York.
Liiketoimintaympäristön muutoksista ks. Nelson, G. Business Intelligence 2.0: Are we there yet? Luettavissa: <http://support.sas.com/resources/papers/proceedings10/040-2010.pdf>.

^{xviii} http://www.lvm.fi/c/document_library/get_file?folderId=3082174&name=DLFE-23659.pdf&title=Julkaisu%2012-2014

^{xix} <http://www.greentechmedia.com/articles/read/meter-data-management-market-to-grow-300-by-2014-according-to-gtm-research/>

^{xx} <http://www.lvm.fi/alyliikenne>

^{xxi} http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation