



Liikenne- ja  
viestintäministeriö

# Big datan hyödyntäminen

## **Liikenne- ja viestintäministeriön**

### **visio**

Hyvinvointia ja kilpailukykyä hyvillä yhteyksillä

### **toiminta-ajatus**

Liikenne- ja viestintäministeriö edistää väestön hyvinvointia ja elinkeinoelämän kilpailukykyä. Huolehdimme toimivista, turvallisista ja edullisista yhteyksistä.

### **arvot**

Rohkeus

Oikeudenmukaisuus

Yhteistyö



Julkaisun nimi

**Big datan hyödyntäminen**

Tekijät

Big datan käyttö -työryhmä, pj. Taru Rastas, siht. Emil Asp

Toimeksiantaja ja asettamispäivämäärä

Liikenne- ja viestintäministeriön big datan käyttö -työryhmä asetettiin 11.12.2013 toimikaudeksi 1.1.2014–30.6.2014

Julkaisusarjan nimi ja numero

**Liikenne- ja viestintäministeriön  
julkaisu 20/2014**

ISSN (verkkojulkaisu) 1795-4045

ISBN (verkkojulkaisu) 978-952-243-407-4

URN [http://urn.fi/URN:ISBN: 978-952-243-407-4](http://urn.fi/URN:ISBN:978-952-243-407-4)

HARE-numero

Asiasanat

big data, tietovarannot, avoin data, tiedon hyödyntäminen

Yhteyshenkilö

Taru Rastas, LVM

Emil Asp, LVM

Tiivistelmä

Erilaisista lähteistä kerättävän ja tallennettavan datan määrä kasvaa voimakkaasti, mutta tietoaineistojen laajamittainen hyödyntäminen on vasta aluillaan. Kehittyneet tiedon keruu-, tallennus- ja käsittelymenetelmät mahdollistavat suurten, eri muodoissa olevien tietomassojen tarkan analyysin. Toiminta ja päätöksenteko voivat muuttua tiedon analysointiin perustuen lähes kaikilla aloilla yhteiskunnassa. Laajojen data-aineistojen sovellusmahdollisuudet ovat lähes rajattomat. Tiedon hyödyntämisestä on tullut yhä vahvemmin kilpailukyvyn edellytys alasta riippumatta.

Liikenne- ja viestintäministeriön big datan käyttö -työryhmän raportissa esitetään luonnos kansalliseksi strategiaksi ja ehdotuksia kansallisiksi strategisiksi toimenpiteiksi, joiden avulla voidaan lisätä suurten tietoaineistojen hyödyntämistä Suomessa. Tietoaineistojen laajemman ja tehokkaamman käytön avulla on saavutettavissa merkittäviä hyötyjä, muun muassa kustannussäästöjä, informoidumpia päätöksiä sekä tarkempia ennusteita toiminnasta. Samalla täytyy huomioida tiedon käsittelyyn ja yhdistelyyn liittyvät riskit, erityisesti yksityisyyden suojan ja eriarvoistumisen osalta.

Strategian tavoitteena on suurten tietoaineistojen laaja ja edistyksellinen, taloudellista kasvua ja yhteiskunnan avoimuutta edistävä käyttö. Lähestymistapana painotetaan tiedon hyödyntämisen monialaisuutta ja -tieteisyyttä sekä tiedon avoimuutta ja yhteistyötä. Raportissa kuvataan laajojen tietoaineistojen käytön kannalta olennaisia toimintaedellytyksiä ja niiden nykytilaa Suomessa. Lisäksi strategiassa tunnistetaan odotetun kehityksen ja innovaatioiden kannalta potentiaalisia tiedon hyödyntämisen ja kehittämisen sovellusalueita.

Työryhmä esittää toimintaedellytyksiin ja Suomen kannalta lupaavimpiin sovellusalueisiin liittyviä toimenpiteitä. Strategialuonnoksen tavoitteen mukaisesti toimet kohdistuvat tiedon laajempaan käyttöön yhtä lailla julkisella sektorilla, elinkeinoelämässä, tutkimuksessa kuin yksittäisten käyttäjien osalta.



Publikation <b>Utnyttjande av big data</b>	
Författare Arbetsgruppen för hantering av big data, ordf. Taru Rastas, sekr. Emil Asp	
Tillsatt av och datum Kommunikationsministeriets arbetsgrupp för hantering av big data tillsattes den 11.12.2013 för mandatperioden 1.1.2014–30.6.2014	
Publikationsseriens namn och nummer <b>Kommunikationsministeriets publikationer 20/2014</b>	ISSN (webbpublikation) 1795-4045 ISBN (webbpublikation) 978-952-243-407-4 URN <a href="http://urn.fi/URN:ISBN:978-952-243-407-4">http://urn.fi/URN:ISBN: 978-952-243-407-4</a> HARE-nummer
Ämnesord  big data, datalager, öppna data, utnyttjande av data	
Kontaktperson Taru Rastas, KM Emil Asp, KM	Rapportens språk finska
Sammandrag Den mängd data som samlas in och lagras från olika källor växer starkt, men än så länge utnyttjas datamängderna inte i särskilt stor skala. Avancerade metoder för insamling, lagring och hantering av data gör det möjligt att noggrant analysera olika former av datamassor. Med hjälp av dataanalyser kan verksamheten och beslutsfattandet omformas inom nästan alla områden i samhället. Möjligheterna att tillämpa mycket stora datamängder (big data) är nästan oändliga. Att utnyttja data har, oavsett bransch, blivit en allt större konkurrensfaktor. Kommunikationsministeriets arbetsgrupp för användning av big data har utarbetat en strategi som siktar på att öka utnyttjandet av stora datamassor i Finland. En bredare och effektivare användning av datamängderna ger betydande fördelar, såsom kostnadsbesparingar, informerade beslut och exaktare verksamhetsprognoser. Samtidigt är det skäl att notera eventuella risker i samband med hantering och kombination av data, särskilt vad gäller integritetsskydd och differentiering (bristande jämlikhet). Målet med strategin är att sporra en omfattande och innovativ användning av mycket stora datamängder och därigenom främja ekonomisk tillväxt och ett öppet samhälle. Fokus ligger på att utnyttja data på ett branschövergripande och tvärvetenskapligt sätt samt på öppna data och samarbete. I strategin beskrivs hurdana verksamhetsförutsättningar som är viktiga vid hantering av mycket stora datamängder och deras nuläge i Finland. Vidare identifieras potentiella tillämpningsområden för att utnyttja och hantera data utifrån förväntade utvecklingstrender och innovationer. Arbetsgruppen föreslår åtgärder med avseende på sådana förutsättningar som påverkar utvecklingen och de tillämpningsområden som är mest lovande för Finland. I enlighet med målet för strategin riktas åtgärderna för att uppmuntra en bredare användning av data såväl till den offentliga sektorn och näringslivet som till forskning och enskilda användare.	



Date  
13 August 2014

Title of publication

**Big data exploitation**

Author(s)

Working group on the use of big data  
chair Taru Rastas, secretary Emil Asp

Commissioned by, date

The Ministry of Transport and Communications' working group on the use of big data was appointed on 11 December 2013 for the period between 1 January and 30 June 2014

Publication series and number

**Ministry of Transport and Communi-  
cations publications 20/2014**

ISSN (online) 1795-4045  
ISBN (online) 978-952-243-407-4  
URN [http://urn.fi/URN:ISBN: 978-952-243-407-4](http://urn.fi/URN:ISBN:978-952-243-407-4)  
Reference number

Keywords

big data, data reserves, open data, exploitation of data

Contact person

Taru Rastas  
Emil Asp

Language of the report

Finnish

Abstract

The amount of data that is collected and recorded from various sources is growing rapidly, but the extensive exploitation of data sets is still in its infancy. Highly developed methods of collecting, recording and processing data will enable an acute analysis of large data masses in various formats. The analysis of data can change operations and decision-making in nearly all sectors of society. The application possibilities of large data sets are nearly unlimited. Regardless of field, the exploitation of data has become an increasingly more important prerequisite for competitiveness.

The report drawn up by the Ministry of Transport and Communications' working group on the use of big data includes a draft national strategy and draft measures to increase the exploitation of large data sets in Finland. More extensive and effective use of data sets will bring significant benefits, such as cost savings, more informed decisions and more accurate forecasts for various operations. At the same time, it is necessary to take into account any risks associated with data processing and combining, particularly those related to privacy protection and inequality.

The goal of the strategy is the extensive and progressive use of large data sets that will promote economic growth and transparency in society. The approach emphasises the diversified and multidisciplinary qualities of data exploitation as well cooperation and the transparency of data. The strategy furthermore describes the operating conditions essential for the use of large data sets and their present status in Finland. The strategy also identifies potential application areas for data exploitation and development with regard to expected development and innovations.

In the report, the working group proposes actions related to the prerequisites for development and the most likely application areas for Finland. In line with the strategic goal, these actions will be targeted at a more extensive use of data both by the public sector, business life, researchers and individual users.

## Sisällysluettelo

1.	Johdanto .....	1
1.1	Raportin tausta ja tavoitteet .....	1
1.2	Raportin rakenne .....	3
1.3	Määrittelyä .....	6
1.4	Perustelut suurten data-aineistojen hyödyntämiselle.....	8
1.5	Big dataan liittyvät haasteet .....	10
1.6	Kansainvälinen kehitys.....	11
1.7	Suomen nykytila ja asema kilpailussa .....	13
2.	Edellytykset .....	14
2.1	Datatietoisuus.....	14
2.2	Koulutus ja osaaminen.....	16
2.3	Tutkimus.....	21
2.4	Infrastrukturi .....	23
2.5	Tekniset käytännöt ja standardit .....	25
2.6	Datan avoimuus ja yhdistely.....	27
2.7	Säätely ja tietosuoja .....	29
2.8	Yhteistyö ja datan vaihto.....	30
2.9	Kokeilut ja T&K -rahoituksen suuntaaminen.....	32
3.	Sovellusalueet ja niiden potentiaali Suomessa.....	33
3.1	Terveys.....	34
3.2	Älykkäät infrastruktuurit .....	35
3.3	Tutkimus.....	37
3.4	Liikenne .....	39
3.5	Teollinen internet .....	40
3.6	Puhdas teknologia .....	42
3.7	Digitaalinen markkinointi ja digitaaliset kuluttajayritykset .....	43
4.	Julkishallinto big datan sovellusalueena .....	45
4.1	Tavoitteena data- ja asiakaslähtöisempi hallinto.....	46
4.2	Suorituskykyä parhailla käytännöillä .....	47
4.3	Kansalaisille yksilöllisempää julkista palvelua.....	48
5.	Sovellusalueet läpileikkaavia teemoja .....	51
5.1	Omadata .....	51
5.2	Itsensä mittaaminen.....	53
5.3	Joukkoistaminen .....	54
5.4	Tiedon etsintä.....	55
5.5	Paikkatieto .....	56
6.	Toimenpiteet .....	57
6.1	Datatietoisuus.....	58
6.2	Koulutus ja osaaminen.....	59
6.3	Tutkimus.....	60
6.4	Infrastrukturi .....	61
6.5	Tekniset käytännöt ja standardit .....	61
6.6	Datan käytettävyys ja avoimuus .....	62
6.7	Säätely .....	63
6.8	Yhteistyö ja datan vaihto.....	63
6.9	Kokeilut ja T&K -rahoituksen suuntaaminen.....	64
6.10	Omadata.....	65
7.	Vaikuttavuus .....	66
8.	Toimeenpano.....	69
	Liitteet .....	70

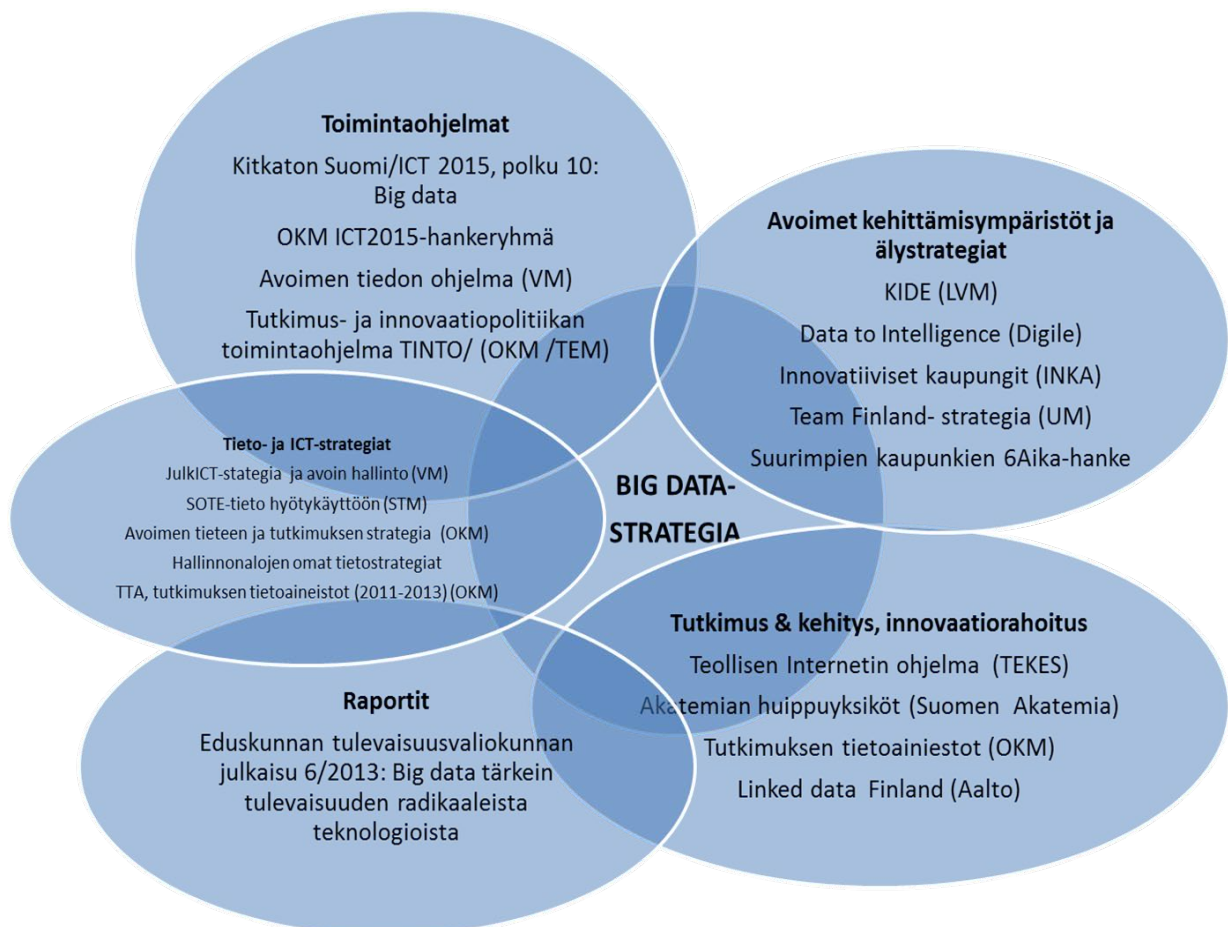
# 1. Johdanto

## 1.1 Raportin tausta ja tavoitteet

Liikenne- ja viestintäministeriö asetti 11.12.2013 big datan käyttö -työryhmän. Työryhmä katsottiin tarpeelliseksi luomaan näkymää ja määrittämään kansallisia kehittämistoimia tietoaineistojen paremmaksi hyödyntämiseksi eri sektoreilla. Useissa eri maissa on laadittu vastaavia strategioita ja EU-tasolla aihe on noussut Digitaaliseen agendalle. Hanke liittyy hallitusohjelmaan kirjattuihin älystrategioihin, joilla ministeriö on KIDE-ohjelman puitteissa edistänyt digitaalisen tiedon hyödyntämistä ja julkisten tietoaineistojen avaamista.

Big data kytkeytyy myös useisiin muihin meneillään oleviin hankkeisiin. Eri hankkeissa ja julkisen hallinnon strategioissa on tärkeä huomioida big data -näkökulma, mikäli sitä ei ole vielä tehty. Samalla hankkeiden keskinäinen koordinointi on olennaista mahdollisimman tehokkaiden tulosten aikaansaamiseksi ja päällekkäisen työn välttämiseksi. Kuvassa 1 on esitetty yhteyksiä työryhmän strategialuonnosta läheisesti sivuaviin kansallisiin hankkeisiin ja strategioihin.

**Kuva 1: Big datan hyödyntämisen strategian yhteydet muihin hankkeisiin**



Työnsä alkuvuodesta 2014 aloittanut työryhmä oli laajapohjainen; mukana on asiantuntijoita niin hallinnosta, tieteen ja tutkimuksen aloilta, innovaatorahoittajista kuin suurista ja pienistä yrityksistä. Liitteessä 1 on esitetty työryhmän kokoonpano. Liitteessä 2 on työryhmän asettamiskirje.

Työryhmän tehtävänä on ollut valmistella ehdotus kansalliseksi big datan hyödyntämisen strategiaksi kansallisina kehittämissinjakoina. Strategialuonnoksessa esitetään kuva laajojen tietoa-aineistojen kehitystilasta Suomessa sekä tunnistetaan potentiaalisia ja Suomen kannalta olennaisia sovellusalueita, joilla tiedon tehokkaammalla ja innovatiivisemmalla käytöllä saataisiin hyötyä niin taloudellisen kasvun kuin yhteiskunnallisen hyvinvoinnin kannalta. Raportin ja esitettyjen toimenpiteiden tavoitteena on kehittää datan laajaa käyttöä koko yhteiskunnassa. Työryhmä esittää tarvittavia edistämistoimia, jotta suurten tietoa-aineistojen käyttö yleistyisi kaikilla sektoreilla, niin yritysmaailmassa, hallinnossa, tutkimuksessa kuin yksittäisten ihmisten kohdalla. Näistä luodaan myöhemmin raportin pohjalta vastuutettu ja priorisoitu toimenpidesuunnitelma. Työryhmän strategiaraportin pohjalta on myös tarkoitettu laatia ehdotus valtioneuvoston periaatepäätökseksi.

Big datan käyttö ei ole uusi asia tutkimuksessa, liike-elämässä tai yhteiskunnassa, mutta laajojen ja eri lähteistä tulevien data-aineistojen hyötykäyttö, analysoinnissa tarvittavat menetelmät sekä osaaminen laajassa mitassa ovat vasta kehittymässä. Tätä arviota kehitysvaiheesta tukevat myös LVM:n teettämä big data -kyselyselvitykseen (2013) osallistuneet organisaatiot, jotka arvioivat aloittavansa alueen kehitysprojekteja vuoden 2016 tienoilla.

Aiheen suomenkielinen termistö ei ole vielä vakiintunutta. Kotimaisten kielten keskuksen (KOTUS) kanssa on tehty yhteistyötä relevantin termistön käyttämiseksi raportissa. Yleisesti käytetyn big data -termin suomalainen vastine on keskuksen mukaan *massadata* tai *iso data*. Liitteessä 3 on listattu käännöksiä tässä raportissa mainituille vierasperäisille sanoille. Työryhmä päätyi raportissa käyttämään englanninkielistä termiä "big data". Tähän ratkaisuun päädyttiin, sillä big data on niin kotimaassa kuin kansainvälisesti alalla jo yleisesti käytössä olevaa kieltä. Toisaalta pelkän suomennetun termin käyttö olisi voinut muun muassa vähentää raportin löydettävyyttä koti- ja ulkomailla.

Ottaen huomioon edellä mainitut datan käyttöön liittyvä piilevä potentiaali, linjausten laaja kohdejoukko ja aiheen yleinen kehitysvaihe, raportin lähestymistavassa korostetaan erityisesti:

- datan mahdollisimman laajaa hyödynnettävyyttä moninlaisiin käyttötarkoituksiin,
- monialaisuutta, poikkihallinnollisuutta ja –tieteellisyyttä tarvittavissa kehitystoimissa, eli laajaa yhteistyötä ja osaamista synergiaetujen saavuttamiseksi,
- datan saatavuutta sekä hyötyjä avoimen datan ja big datan yhdistelystä,
- datamassojen käsittelyssä tarvittavien uusien menetelmien, teknologioiden, toimintatapojen ja osaamisen kehitystarpeita,
- datan hyödyntämisestä saatavaa taloudellista ja ei-taloudellista arvoa

Uusien big data -menetelmien tuomat mahdollisuudet ja hyödyt vaativat usein monialaisuutta ja poikkihallinnollista sekä –tieteellistä lähestymistapaa. Raportin lähtökohtana



nähdään tarve laajalle yhteistyölle. Data-aineistot, joiden yhdistelystä ja soveltamisesta syntyy arvoa, ovat usein organisaatioiden ja instituutioiden omistuksessa. Laajaa yhteistyötä tarvitaan siten niin hallinnon, yritysten kuin esimerkiksi järjestöjen ja kansalaisten kesken.

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 raportin painotuksista koskeekin avointa dataa. Tietoa tulisi olla saatavilla ja käytettävissä mahdollisimman avoimesti, toki esimerkiksi yksityisyyden suojaa vaarantamatta.

Teknologioiden, datan keruu- ja analyysimenetelmien sekä anturien kehittyessä datasta tehtävät havainnot ja yleisesti koko big data -kehitys on vielä murrosvaiheessa. Kehittyviä uusia ratkaisuja on vaikea ennustaa, mutta aikainen kehitysvaihe antaa myös suomalaisille toimijoille mahdollisuuden kehittää alan teknologiaa ja palveluja.

Strategiaehdotuksen ajatuksena on löytää Suomen kannalta oleellisia alueita, joilla kerätävä tieto, uudet tiedonkäsittelymenetelmät ja laajempi tiedon hyödyntäminen voivat luoda arvoa. Työn tavoitteet ja toimet kohdistuvat siten niin yksityiseen kuin julkiseen sektoriin sekä käyttäjälähtöisen datan hallinnan kehittämiseen.

**Raportin ja siinä esitettyjen kansallisten toimenpide-ehdotusten tavoitteena on isojen datamassojen ja uusien big data -menetelmien laajempi hyödyntäminen:**

- **yritysten liiketoiminnan kasvattamisessa, kilpailukyvyssä ja uuden liiketoiminnan synnyttämisessä,**
- **julkisen hallinnon ja sen palvelujen kehittämässä ja yhteentoimivuudessa,**
- **tutkimuksen laadun ja yhteiskunnallisen vaikuttavuuden parantamisessa sekä**
- **yksilön mahdollisuuksissa hyödyntää omia tietojaan.**

## 1.2 Raportin rakenne

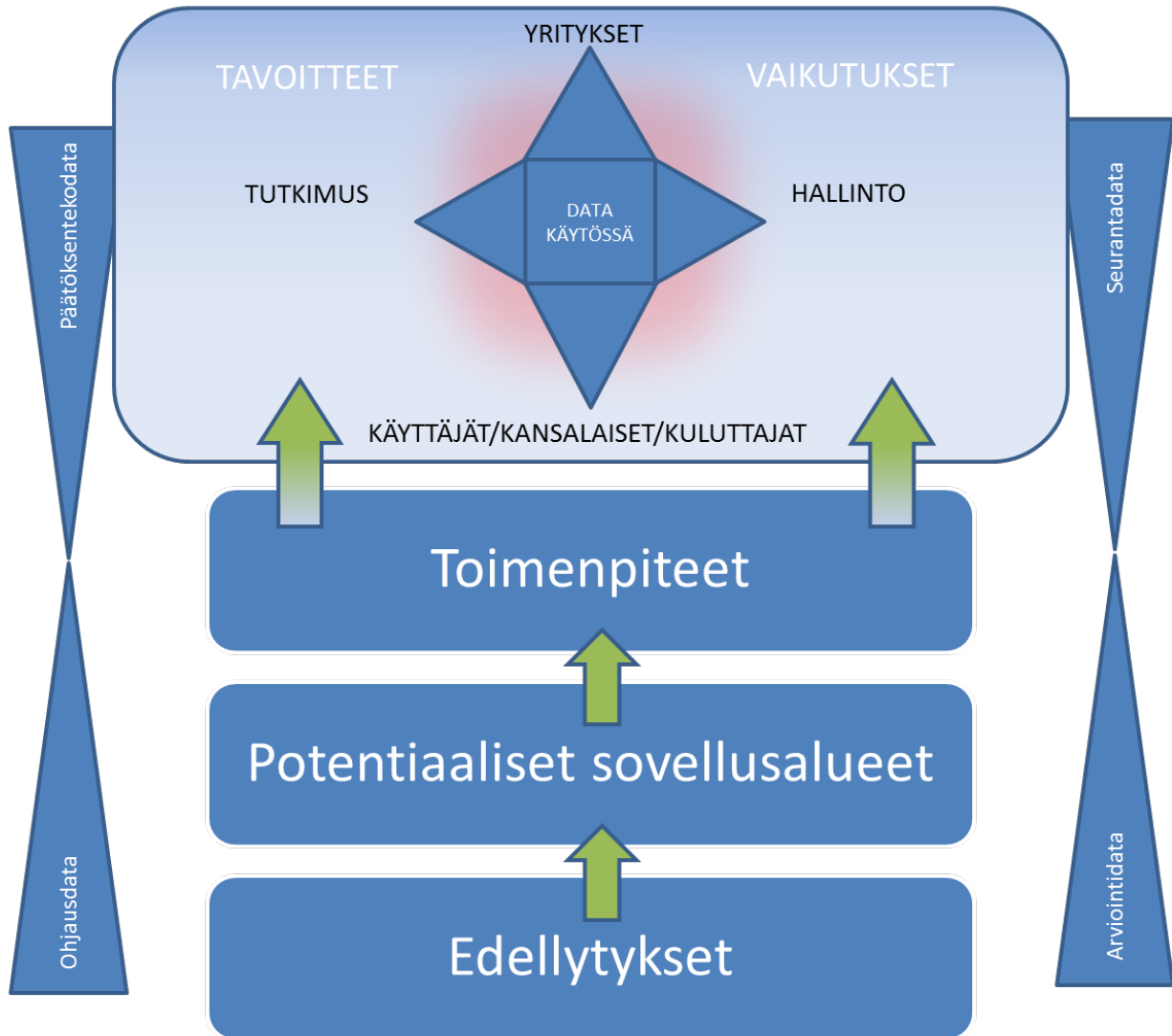
Oheisessa kuvassa 2 on esitetty tämän raportin rakenne. Sen perustana luvussa kaksi esitetään isojen datamassojen hyödyntämiskehitykseen vaikuttavat edellytykset sekä niiden nykytila ja kehitystarpeita. Luvussa kolme tuodaan esille tunnistetut, Suomelle potentiaaliset sovellusalueet sekä luvussa viisi läpileikkaavia teemoja. Luvussa neljä on käyty läpi julkishallintoa omana sovellusalueenaan.

Edellytyksistä ja sovellusalueista johdetut toimenpide-ehdotukset ovat koottuna luvussa kuusi. Luvussa seitsemän arvioidaan toimien vaikuttavuutta ja aiheen merkitystä eri toimijoiden näkökulmasta. Raportin tavoitteena on siten hakea vaikuttavuutta esitettyjen toimien ja eri käyttökohteissa datan laajan hyötykäytön kautta.

Raportti itsessään voidaan kuvata sekä datalähtöisenä toimintana että myös sen toteutuksessa ja seurannassa tarvittavana datana. Edellytyksistä saatava data sekä soveltamisessa syntyvä data ohjaavat toimenpiteisiin liittyvää päätöksentekoa. Datana hyöty-

käyttöä eri sektoreilla tulee myös jatkuvasti seurata. Lisäksi tulee arvioida edellytysten ja sovellusalueiden kehittymistä.

**Kuva 2. Raportin rakenne**



Raportissa esitetään big datan hyödyntämiseen liittyvän toimintaympäristön kuvaus, jollaista ei Suomessa aiemmin ole tehty. Se keskittyy tunnistamaan, kuvaamaan ja luomaan edellytyksiä, joita datan ja muun muassa uusien analyysimenetelmien ja -ratkaisuiden laajassa hyödyntämisessä tarvitaan.

Työryhmä tunnisti ja priorisoi työssään Suomen kannalta potentiaalisia suurten tietoi-  
neistojen hyödyntämisen sovellusalueita. Tunnistetuilla aloilla meneillään olevaa kehitys-  
tä tulisi seurata erityisen tarkasti ja kohdistaa sovellusalueisiin esitettyjä toimenpiteitä  
esimerkiksi kokeilujen, tutkimusrahoituksen tai kehitysalustojen muodossa. Taulukossa 1  
on esitelty yhteenveto ehdotetuista toimista ja toimenpiteiden kannalta olennaisista toi-  
mijaryhmistä. Toimenpiteiden kannalta tärkeitä esimerkkiorganisaatioita on kirjattu  
muun muassa eri organisaatioilta lausuntokierroksen yhteydessä saadun palautteen ja  
tiedon perusteella. Lista ei ole täysin kattava, vaan tarkentuu ja täydentyy toimenpide-  
suunnitelman laatimisen yhteydessä. Saaduista 68 lausunnosta on myös poimittu sitaat-  
teja tekstiin.

Taulukko 1: Yhteenvedo toimenpide-ehdotuksista ja keskeisiä toimijoita

TOIMENPIDE-ALUE	TOIMENPIDE-EHDOTUKSET				
<b>Datatietoisuus</b>	1. Datalähettiläät	2. Klusterin luominen	3. Datavastaavat, käytänteet		
<i>Keskeisiä toimijoita:</i>	<i>Ministeriöt, yritykset, tutkimuslaitokset (esim. KTK), oppilaitokset (esim. yliopistot ja ammattikorkeakoulut), yhdistykset (esim. Tieke), järjestöt (esim. TEK, TTL, Ohjelmistoyrittäjät, Teknologiateollisuus), Keskuskauppakamari, tietosuojavaltuutettu, SHOKIt</i>				
<b>Koulutus ja osaaminen</b>	1. Datan käyttöön liittyvä opetus	2. Koulutustarjonta-kartoitus	3. Osaamisprofiilit ja koulutusohjelmien kehittäminen	4. Koulutusyhteistyö verkosto (EU)	5. Opiskeluympäristöjen kehittäminen
	6. Big data -harjoittelut	7. Täydennyskoulutus	8. Yritysten omien big data -hautomoiden tukeminen		
	<i>OKM, Opetushallitus, oppilaitokset, koulutusorganisaatiot, yritykset, liitot ja järjestöt (esim. SAK), kaupungit (esim. Tampere), tietosuojavaltuutettu, järjestöt (esim. Teknologiateollisuus)</i>				
<b>Tutkimus</b>	1. Tutkimusrahoituksen kanavointi big data -tutkimukseen	2. Tutkimusinfra-riittävyys- ja tulevaisuusarvioinnit	3. Tutkimustiedon aineistojen saatavuuden parantaminen	4. Datan jakamiseen kannustaminen rahoituksen ohjauksella	5. Globaali yhteistyö tutkimustiedon jakamisessa
	<i>OKM, tutkimus- ja innovaationeuvosto, oppilaitokset, Suomen Akatemia, TEKES, CSC, SHOKIt, tutkimuslaitokset (esim. VTT, HIIT, KTK), virastot (MML, IL, THL, SYKE), tietosuojavaltuutettu, liitot ja järjestöt, yritykset</i>				
<b>Infrastruktuuri</b>	1. Infra kehittäminen verkostoyhteistyöllä, sis. tutkimus/koulutus	2. Huippuluokan globaalit ICT-yhteydet	3. Kokeilu- ja testausympäristöjen tukeminen		
	<i>LVM, TEM, Viestintävirasto, TEKES, kaupungit (esim. Tampere) ja kunnat, yritykset, oppilaitokset, järjestöt, tietosuojavaltuutettu, liitot ja järjestöt, SHOKIt, CSC</i>				
<b>Tekniset käytännöt ja standardit</b>	1. Standardien tunnistaminen ja standardointiyhteistyö	2. Läpinäkyvän ja turvallisen henkilötiedon käyttämisen infrakehitys			
	<i>Standardointiliitot- ja yhdistykset, yritykset, yhdistykset, kunnat/kaupungit, TEKES, tietosuojavaltuutettu</i>				
<b>Datan käytettävyys ja avoimuus</b>	1. Big datan hyödyntämisen kehittäjä hallintoon	2. Yhteishankkeet kunta- ja kaupunkiympäristössä	3. Kansallisen dataporttiin edelleen kehittämisen big data huomioiden	4. Virastojen hyödyllisten aineistojen tunnistaminen ja pilotoinnit	
	<i>VM ja muut ministeriöt, virastot, kunnat ja kaupungit, kansalaisjärjestöt, liitot ja järjestöt, yritykset, oppilaitokset, tietosuojavaltuutettu</i>				
<b>Säätely</b>	1. Datan hyödyntämisen näkökulma lainvalmistelussa	2. Säätelyn, kehityksen ja yksityisyyden suojan tasapainottaminen	3. Yhteistyönä rakennettavat käytänteet		
	<i>OM, STM, OKM, SM, VM ja muut ministeriöt, alan toimijat (yritykset/järjestöt/liitot), tietosuojavaltuutettu</i>				
<b>Yhteistyö ja datan vaihto</b>	1. Datahub	2. PK- ja start-up-yritysten pääsy tietoa-ineistoresusseihin	3. Oppilaitos-/tutkimuslaitosyhteistyö ja teknologiasirto	4. Osallistuminen EU:n Big /NESSI – datafoorumeihin	5. Joukkoistamisen hyödyntämismallien kehittäminen
	<i>LVM, TEM, yritykset (esim. IBM, Fortum, CGI), liitot ja järjestöt (esim. TTL, Tekniikan Akateemiset, Teknologiateollisuus, Energiateollisuus), yhdistykset (esim. Tieke, Start-up-yhdistykset), Tekes, Sitra, oppilaitokset (esim. HY, JY, TY, Metropolia, Savonia), tutkimusorganisaatiot, kunnat ja kaupungit (esim. Helsinki, Tampere), tietosuojavaltuutettu</i>				
<b>T&amp;K / Rahoitus</b>	1. Kansalliset/EU-rahoitusohjelmat big data -painotuksella	2. Kokeilujen tukeminen ja saatujen kokemusten jakaminen	3. Kokeilevat hankkeet kaupallistamisen ja tulojaon näkökulmista	4. Pääomarahituksen saatavuuden edistäminen	
	<i>TEM, TEKES, Horisontti 2020, Akatemia, SHOKIt, yritykset, tutkimuslaitokset, + oppilaitokset, Elyt, kaupungit/kunnat, liitot ja järjestöt (esim. Proliitto)</i>				
<b>Omadata</b>	1. Yritysyhteistyö omadata-mallien testauksessa	2. Omadata-alustan luominen	3. Soveltuvuusarvioinnit omadatan käyttämistä viranomaistoiminnassa	4. Verkostoyhteistyö muiden omadata-toimijamaiden kanssa	
	<i>LVM, VM, TEM, yritykset, TEKES, järjestöt (OKF), eri alojen yritykset, kaupungit/kunnat (esim. Helsinki, Tampere), tietosuojavaltuutettu, oppilaitokset, tutkimuslaitokset (esim. KTK)</i>				
<b>Vaikutavuus</b>	1. Taloudellinen ja yhteiskunnallinen analyysi toimenpiteiden vaikutavuudesta	2. Vaikutavuusanalyysi avoimen datan ohjelman kanssa			
	<i>VM, LVM, tutkimuslaitokset, oppilaitokset, toimialajärjestöt, yritykset, tietosuojavaltuutettu, liitot ja järjestöt (esim. FiCom, Teknologiateollisuus, SAK jne.)</i>				
<b>Toimeenpano</b>	1. Toimeenpanosuunnitelma vastuuneen, määrärahatarpeiden arviointi				
	<i>LVM valmisteleeh yhdessä toimijoiden kanssa, päätöksenteko VN/taipo (vn pp)</i>				

Suurten data-aineistojen hyödyntämisen vaikutukset näkyvät kaikilla aloilla, niin yritysmaailmassa, hallinnossa, tutkimuksessa, kuin yksittäisten ihmisten kohdalla. On tärkeää huomata, että nämä vaikutukset heijastelevat myös takaisinpäin: vaikuttavuuden arvioinnin ja mittareiden kehityksen kautta tarvittavia toimia ja edellytyksiä voidaan muokata jatkuvasti. Tätä datalähtöisen toiminnan periaatetta sovelletaan myös raportin rakenteena ja sisältönä.

Nopeasti kehittyvässä ja muuttuvassa toimintakentässä perinteinen lähestymistapa ja pitkäaikaisen, esimerkiksi kymmenen vuoden päähän ulottuvan tarkan vision luominen ei siis välttämättä ole optimaalisin tapa toimia. Näin ollen strategioissa usein esitettyä määrämutoista kansallista visiota ei tässä raportissa ole eritelty. Datan hyödyntäminen on vasta alussa<sup>1 2</sup>. IDC:n mukaan vuonna 2013 kaikesta mahdollisesta analysoitavasta datasta analysoitiin vain viisi prosenttia.

### 1.3 Määrittelyä

44 00 000 000 000 000 000 000 000 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. Erilaisten laitteiden, palveluiden ja anturijärjestelmien yleistymisen ja lisääntyvän käytön seurauksena digitaalinen avaruus kasvaa vuosittain 40 %:lla eli kymmenkertaiseksi vuoteen 2020 mennessä.<sup>3</sup>

Mistä dataa oikein syntyy sellaisia määriä, niin nopeasti tai niin monimuotoisena, että voidaan puhua big datasta? Pääasiassa verkkoon kytketyistä laitteista, niiden lähettämästä datasta, erilaista anturijärjestelmistä, sosiaalisesta mediasta, erilaisista verkon yli tehtävistä transaktioista, perinteisestä mediasta, tietojärjestelmistä, yritysten liiketoimintaan liittyvistä ohjaus- ja raportointijärjestelmistä, erilaisista digitaalisista arkistoista ja rekistereistä<sup>4</sup>. Dataa syntyy kiihtyvällä vauhdilla lähes kaikesta internetiin jollakin tavalla liitetystä asiasta ja järjestelmästä.

Syntyvä data on muodoltaan hyvinkin monimuotoista, aina yksinkertaisista tekstidokumenteista ja sosiaalisen median päivityksistä reaaliaikaisen videovirran ja verkkoon kytketyn hissien anturitiedon kautta monimutkaisten satelliittijärjestelmien tuottamaan dataan.

Termille 'big data' ei ole yhteisesti jaettua ja yksiselitteistä määrittelyä<sup>5</sup>. Useimmiten käytetään kolmen V:n määritelmää, jolla viitataan sekä datan määrään (volume), syntyvauhtiin (velocity), sekä datan monimuotoisuuteen ja strukturoimattomuuteen (variety). Eri näkökulmista, muun muassa eri palveluntarjoajien määritelmässä, painotetaan ilmiön eri puolia. Suoritinvalmistaja Intel määrittelee big datan puhtaasti karttuvan tietoaineis-

<sup>1</sup> Deloitte. Data: A Growing Problem. [http://www.deloitte.com/assets/Dcom-UnitedStates/Local%20Assets/Documents/IMOs/Deloitte%20Analytics/us\\_ba\\_BigDataSalesSheet\\_10292012.pdf](http://www.deloitte.com/assets/Dcom-UnitedStates/Local%20Assets/Documents/IMOs/Deloitte%20Analytics/us_ba_BigDataSalesSheet_10292012.pdf)

<sup>2</sup> Gantz & Reinsel (2012), IDC iView: THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. <http://idcdocserv.com/1414>

<sup>3</sup> Turner, Gantz, Reinsel & Minton (2014), The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things. <http://idcdocserv.com/1678>

<sup>4</sup> Kapow Software. Intelligence by Variety - Where to Find and Access Big Data.

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

<sup>5</sup> Ward & Barker (2013), Undefined By Data: A Survey of Big Data Definitions <http://arxiv.org/abs/1309.5821>

ton koon mukaan (mediaaniarvona 300 teratavua dataa viikossa), ohjelmistoyritys Microsoft lähestyy asiaa analyysin vaadittavien monimutkaisten menetelmien näkökulmasta. Lisäksi termiä voidaan lähestyä myös analyysiin käytettävien työkalujen kautta, muun muassa Hadoopin ja NoSqlin.

Big datan voi käsittää myös tietynlaisena tiedon paradigman muutoksena. Sen myötä yrityksissä ja hallinnossa voidaan yhä useammin tehdä päätöksiä suoraan kerättyyn ja mitattuun tietoon perustuen. Tutkimuksessa on mahdollista muodostaa teoriaa uusilla tavoilla, kun valtavien tietomassojen analysointi ja yhdistely on mahdollista, joskin suuriin tietoaisteistoihin perustuvaa tutkimusta on tehty jo aiemmin. Big data voidaan ymmärtää myös kokeellista, teoreettista ja laskennallista tiedeparadigmia yhdistävänä uudenlaisena paradigmana.

#### **Big datan määrittelyissä käytettyjä asioita:**

- Koko, varastoitavan tai käytettävän datan ja aineistojen suuri määrä.
- Datan ja datavirran nopeus, tähän liittyy usein reaaliaikaisuus.
- Monimuotoisuus, ei-rakenteisuus ja sijainti. Käytettävien tietoaisteistojen rakenne ja sijainti vaihtelee ja vaatii useimmiten yhdistelyä.
- Käsittelyyn tarvittavat teknologiat.
- Datan analyysistä saatava hyöty tai datan arvo hyödykkeenä.

Terminä big data on siis suhteellisen löyhästi määritetty, osan määrittelijöistä korostaessa nimenomaan kokoa, osan pelkkiä menetelmiä, ja osan painottaessa big datasta seuraavia muutoksia.

Joka tapauksessa big datan hyödyntämisen yhtenä kynnyksenä on tiedon käsitteilyyn ja analysointiin vaadittavien uusien menetelmien tarve. Perinteiset menetelmät eivät big data-analyysissä ole riittäviä.

#### **LVM:n teettämässä Big data Suomessa -raportissa pohdittiin datan hyödyntämisen eri ulottuvuuksia:**

" - - puhutaan siis datasta, jota on paljon, jota tulee nopeasti lisää ja joka on muodoltaan vaihtelevaa. Syntyhetkellä sen oikeellisuus, oleellisuus ja arvo ovat konteksti- ja aikasidonnaisia. Data voi olla arvotonta nyt, mutta arvokasta tulevaisuudessa. Yhdelle organisaatiolle kullannarvoinen data on toiselle arvotonta. Ratkaisut painivat datan tallentamiseen, yhdistelemiseen, siirtämiseen ja ennen kaikkea analysointiin eli hyödyntämiseen liittyvien ongelmien kanssa. Data itsessään ei missään mitta-kaavassa tai muodossa ole arvokasta, vaan vasta sen organisointi ja analysointi tekee siitä arvokasta."

**Tässä raportissa on big dataan keskitytty nimenomaan datan laajemman hyödyntämisen näkökulmasta. Big dataksi käsitetään laajasti ainakin osan yllämainitusta tunnusmerkeistä täyttävä datan ja tietoaisteistojen käyttö sekä analytiikka.**

Lähde: LVM 2013, Big data Suomessa: [http://www.lvm.fi/docs/fi/2497123\\_DLFE-21601.pdf](http://www.lvm.fi/docs/fi/2497123_DLFE-21601.pdf)

Raportin laaja-alaisuudesta johtuen termiä "big data" ei määritellä tiukasti esimerkiksi pelkän tiedon määrän, teknologisten ratkaisujen tai käsittelymenetelmien kautta.

## 1.4 Perustelut suurten data-aineistojen hyödyntämiselle

Teknologisen kehityksen ja etenkin pilvipalveluiden kehityksen myötä digitaalisen tiedon tallennus- ja prosessointikustannukset ovat laskeneet jopa murto-osiin aiempiin kustannuksiin verrattuna. Nämä kehitystrendit yhdessä uusien digitaalista tietoa käsittelevien laskentamenetelmien kehityksen kanssa ovat tuoneet suurten data-aineistojen hyödyntämismahdollisuudet myös muiden kuin tutkimuslaitosten ja globaalien suuryritysten ulottuville. Hyödyntämismahdollisuuksien etsintää ja kehitystä on vauhdittanut vielä tieto- ja viestintäteknologioiden sekä viestintäverkkojen muassaan tuoma mahdollisuus tiedon reaaliaikaiseen käsittelyyn ja siihen liittyvän päätöksenteon – yhä useammin automaattisen päätöksenteon – reaaliaikaisuuteen. Teknologiset innovaatiot ovat mahdollistaneet tiedon keräämisen, käytön ja sen pohjalta toimintamallien muovaamisen taloudellisesti.

Big datan laajemmalla hyödyntämisellä haetaan useita selkeitä etuja niin yritysmaailmassa kuin julkisessa hallinnossa. Näihin mahdollisuuksiin ja hyötyihin kuuluvat muun muassa toiminnan optimointi ja siitä seuraavat säästöt, tarkemman tiedon saaminen päätöksenteon tueksi, tarkemman tilannekuvan saaminen, uusien mallien ja yhteyksien löytäminen mallintamalla, parempi asiakaspalvelu sekä tulevan ennustaminen.

Yritysten liiketoiminnalle on tietoaineistojen hyödyntämisestä tullut selkeä kilpailuvaltti<sup>6</sup> <sup>7</sup>. Yhä useampi yritys ottaa käyttöönsä suurten tietomassojen analyysiin perustuvia ratkaisuja<sup>8</sup>, joiden avulla saatava informaatio vaikuttaa suoraan yrityksen toimintaan. Tiedon yhä paremmasta hyödyntämisestä on tulossa ehdoton menestymisen edellytys<sup>9</sup>. Uusien tiedon hyödyntämiseen keskittyvien ratkaisujen käyttöönoton merkityksen ja kiihtyvän vauhdin voi havaita myös markkinatutkimusyhtiö IDC:n laskelmista. Ne ennustavat big data -teknologia- ja palvelumarkkinoiden kasvavan jopa 27 %:n vuositahtilla. IDC:n mukaan näiden markkinoiden koko vuonna 2017 tulee olemaan jopa 32,4 miljardia dollaria.<sup>10</sup> Eri tutkimus- ja konsulttiyhtiöiden selvityksissä ja tutkimuksissa nähdään big data -kehitys vahvasti koko taloutta ja yhteiskuntaa muovaavana tekijänä<sup>11</sup>.

<sup>6</sup> EMC. Gain Competitive Advantage From Big Data. <http://www.emc.com/collateral/emc-perspective/h8729-gain-comp-advntng-ep.pdf>

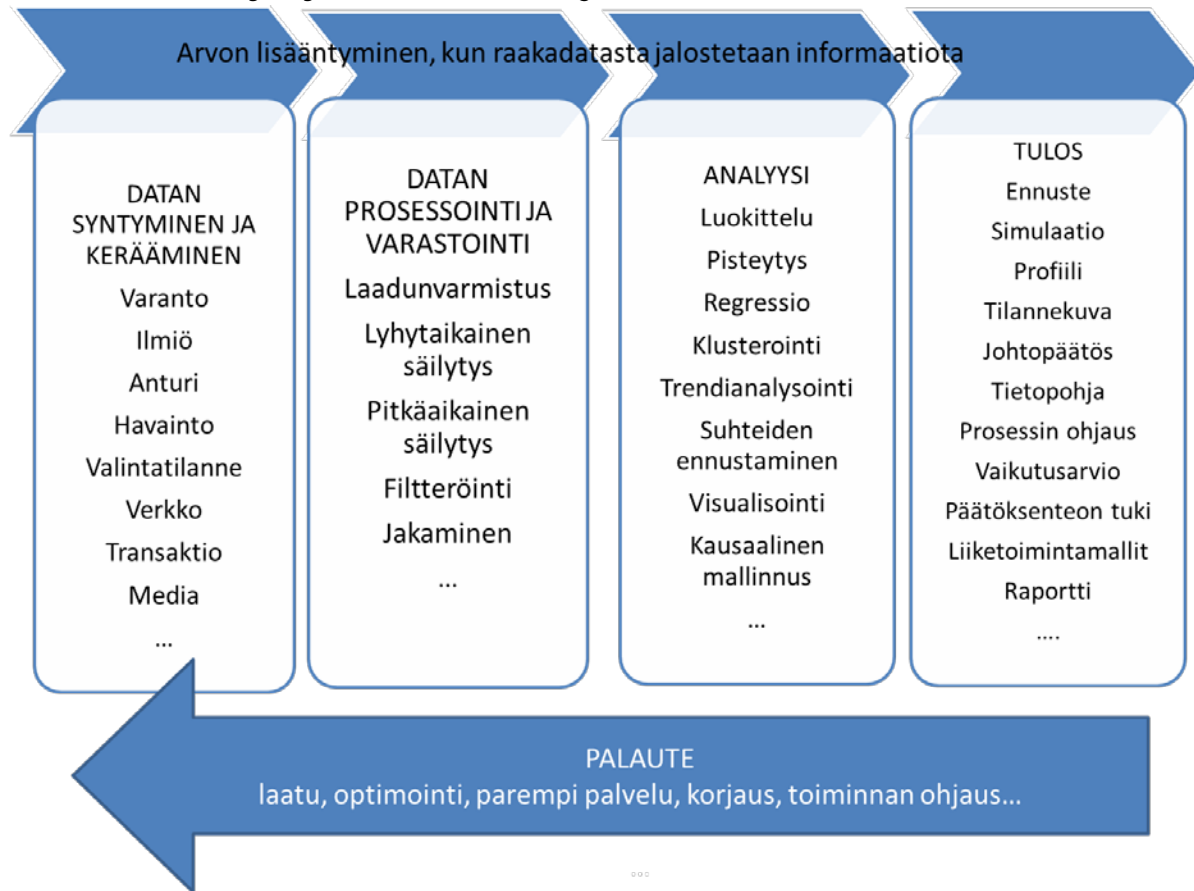
<sup>7</sup> Davenport & Dyché (2013), Big Data in Big Companies. <http://www.sas.com/resources/asset/Big-Data-in-Big-Companies.pdf>

<sup>8</sup> TCS. Big Data Study – The 10 Key Findings. <http://sites.tcs.com/big-data-study/big-data-study-key-findings/>

<sup>9</sup> The Economist Intelligence Unit (2012), The Deciding Factor: Big data and decision-making. <http://www.capgemini.com/resources/the-deciding-factor-big-data-decision-making>

<sup>10</sup> IDC. Big Data and Analytics – An IDC Four Pillar Research Area. <http://www.idc.com/prodserv/FourPillars/bigData/index.jsp>

<sup>11</sup> McKinsey Global Institute (2013), Disruptive technologies: Advances that will transform life, business, and the global economy. [http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Disruptive%20technologies/MGI\\_Disruptive\\_technologies\\_Full\\_report\\_May2013.ashx](http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Disruptive%20technologies/MGI_Disruptive_technologies_Full_report_May2013.ashx)

**Kuva 3: Datan hyödyntämisen arvoketju**

Kerättävää tai käytettävää dataa analysoimalla haetaan monipuolista ymmärrystä ilmiöistä, josta dataa on kerätty. Analyysien tavoitteina voivat olla muun muassa tuki päätöksille, erilaisten profiilien luonti, erilaiset simulaatiot ja prosessien ohjaukseen vaikuttavat tekijät. Raakadatasta luodaan käsittelyn ja analyysin jälkeen jalostuneempaa informaatiota, jonka muotoja on esitetty datan hyödyntämisen arvoketjua esittävän kuvan 3 TULOS-laatikossa.

Tulosten pohjalta pyritään vaikuttamaan suoraan organisaation toimintaan eri tavoilla, esimerkiksi parantamalla palvelun laatua, korjaamalla toimintaa, tai yleisesti optimoimalla. Yllä olevan kuvan yksinkertaistettu arvoketju kuvaa pääpiirteissään sen, miten datalla tavoitellaan ja saavutetaan hyötyjä. Arvoketjun ja hyödyntämisen eri vaiheiden välillä on myös runsaasti takaisinkytkentöjä. Datan jalostaminen tiedoksi ja tämän tiedon merkityksen ymmärtäminen ei ole aina kuvatun kaltaisesti täysin suoraviivaista.

Usein dataa varastoidaan myös myöhempää, ei vielä ennakolta määriteltyä käyttöä varten. Toisaalta datan arvoketju saattaa joissain tapauksessa päättyä myös esimerkiksi arkaluontoisen tai säilyttämiseen kelpaamattoman datan hävittämiseen.

Datan arvoketjun eri vaiheisiin syntyy myös uusia, big dataan liittyviä palveluita tarjoavia toimijoita. Palvelut voivat liittyä niin datan keruuseen, laadunvarmistukseen, prosessointiin kuin analyysiin. Myös datan vaihtoon ja kauppaamiseen liittyviä uusia palveluita voi syntyä. Uudenlainen dataekosysteemi tarjoaa mahdollisuuksia muun muassa tiettyihin alueisiin erikoistuneille pk-yrityksille.

### Uutta tietointensiivinen palveluliiketoimintaa perinteisellä alalla: case Eniram

Suomalainen Eniram kehittää merenkulkualalle polttoaineenkulutusta vähentäviä analyysijärjestelmiä. Järjestelmät keräävät jatkuvasti reaaliaikaista dataa, ja esittävät avainmittarit visualisoituina graafiseen muotoon. Järjestelmä kehittää automaattisesti toimintaansa koneellisen oppimisen kautta, ja säätää ohjausjärjestelmien toimintaa vertaamalla olosuhdetietoja anturien keräämään aikaisempaan tietoon. Monimuuttujainen analyysikehikko laskee jatkuvasti optimaalisinta säätöä suhteessa aluksen käytäytymiseen vaikuttaviin voimiin.

Minimoitu vesivastuksen määrä, alentuneet polttoainekustannukset ja pienentyneet päästöt ovat suoria tuloksia suuren tietomassan soveltamisesta liiketoiminnallisesti merkittävien säästöjen aikaansaamiseksi. Yhdessä polttoaineen ympäristöystävällisyys- ja energiatehokkuus-vaatimusten kanssa dynaamisten tietomassojen ja –varantojen hyödyntäminen ja tuotteistaminen on tuonut yritykselle kilpailullisen etulyöntiaseman.

Lähteet: <http://www.thedigitalship.com/home/item/314-fuel-saving-software-released-by-eniram/314-fuel-saving-software-released-by-eniram> , [www.eniram.fi](http://www.eniram.fi)

## 1.5 Big dataan liittyvät haasteet

Toisaalta big datan aikakausi tuo mukanaan myös erilaisia uhkakuvia. Ylipäänsä tulee huomioida myös big datan käyttöön liittyvät mahdolliset riskit, esimerkiksi yksityisyyden suojaan liittyen. Kehityksessä ja erilaisissa toimenpiteissä täytyy pohtia myös mahdollisia ongelmia.

Edellä mainittuihin ongelmiin kuuluvat muun muassa yksityisyyden ja luottamuksen vaarantuminen, (usein henkilöitä koskevan) tiedon omistajuuden keskittyminen vain harvoille toimijoille, liiallinen käyttäjien seuraaminen ja liiallisesta profiloinnista mahdollisesti seuraava diskriminointi, esimerkiksi erilaisten maksujen suhteeton kallistuminen profiiliin perustuen.<sup>12</sup>

Henkilötietoriskit voivat yleisellä tasolla liittyä paitsi suppeaan tai laajamittaiseen tietojen väärinkäyttöön, esimerkiksi virheisiin tai taitamattomuuteen suurten tietoaisteiden käsittelyssä, analysoinnissa ja tulosten tulkinnassa. Riskien merkitys voi olla sitä suurempi, mitä merkittävämmistä tietoaisteista on kyse ja kuinka kriittistä analyysien tulosten hyödyntäminen toiminnan kannalta on.

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 ymmärtämättömästi voi johtaa myös täysin virheellisiin kategoriainteihin. Tulkinnassa on myös muistettava, että data voi olla valikoituneesti painottunutta, varsinkin kun kyse on henkilöistä kerättävästä datasta. Tietomäärät ja big datan uusien mahdollisuuksien vastuullinen hyödyntäminen korostaa tiedon tulkinnan merkitys-

<sup>12</sup> European data protection Supervisor: Privacy and competitiveness in the age of big data (2014), [https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2014/14-03-26\\_competition\\_law\\_big\\_data\\_EN.pdf](https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2014/14-03-26_competition_law_big_data_EN.pdf)



tä. Tiedon hyödyntäminen vaatii hyvää ymmärrystä esimerkiksi tilastollisesta päättelystä<sup>13</sup> ja eettisestä päätöksenteosta.

Yleisesti datan luotettavuus ja oikeellisuus on kriittinen kysymys. Tehtävien analyysien laatu riippuu datan laadusta. Erilaisilla organisaatioilla ja toimijoilla on myös erilaisia vaatimuksia datan laadun suhteen. Joihinkin käyttötarkoituksiin riittää karkeamman tarkkuustason data, joissain tapauksissa taas datalta vaaditaan ehdotonta tarkkuutta ja luotettavuutta. Oleellista on se, että datan analysoijat tietävät tietojensa laatutason.

Lähes kaiken kytkeytyessä verkkoon, myös tietoturvallisuus sekä ylipäänsä luottamus nousevat ylipäänsä aivan keskeiseksi asiaksi. Organisaatioiden ja samalla koko yhteiskunnan toimivuus on pitkälti kiinni luotettavasta tietoturvasta. Uusien tietoekosysteemien syntyessä myös keskinäiset riippuvuudet kasvavat.

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. Ulkopuolisen analyysisovelluksen tulee siis päästä suoraan järjestelmään. Niinpä eteen tulee tilanteita, joissa tietoa hallussa pitävä organisaatio joutuu päättämään, luottaako se ulkopuoliseen sovellukseen niin paljon, että päästää sen suoraan sisään järjestelmänsä ja käsiksi tietoineistoihinsa. Erilaisia käytäntöjä täytyy miettiä tietoturvan näkökulmasta tarkasti.

## 1.6 Kansainvälinen kehitys

Euroopan ja Suomen kehityksen katsotaan yleisesti olevan muutaman vuoden jäljessä kansainvälisesti suurista big datan hyödyntäjämaita, erityisesti USA:ta. Eri maiden tietostrategioita tarkastellessa huomaa, että big data nähdään yhtenä oleellisimmista kehityskuluista säästöjen ja kilpailukyvyn kannalta. Tietomassojen laajaa hyödyntämistä pidetään oleellisena uuden kasvun lähteenä.

Monissa maissa panostetaan mittavasti big datan alueelle. Big data -maailmassa kilpailu on kovaa ja investoinnit alalle suuria. Suomi on mukana globaalissa datakilpailussa. Data, siitä jalostettava tieto ja ymmärrys eivät tunne maantieteellisiä rajoja perinteisten fyysisten hyödykkeiden tavoin.

Erilaiset big data -alan yritykset kasvavat nopeasti ja keräävät mittavia sijoituspääomia kehityksen turvaksi<sup>14</sup>. Investoinnit big dataan ja analytiikkaan ovat vahvassa kasvussa<sup>15</sup>. Vuonna 2013 big data -alalla toimiviin yrityksiin investoitiin miljardeja, suurimpien yksittäisten investointien ollessa satojen miljoonien suuruisia<sup>16</sup>.

IDC:n mukaan suuryritykset näkevät big datassa valtavan potentiaalin ja ovat lähiaikoina investoimassa omaan big data -kehitykseensä huomattavia summia, mikäli eivät näin ole jo tehneet. IDC:n kyselytutkimukseen vastanneista suuryrityksistä 74 % näki big datan

<sup>13</sup> Kts. big dataan liittyviä kriittisiä huomioita esim. Boyd, [http://www.tandfonline.com/doi/abs/10.1080/U5hTvz1\\_t8E](http://www.tandfonline.com/doi/abs/10.1080/U5hTvz1_t8E)

<sup>14</sup> <http://www.experfy.com/blog/surge-in-big-data-venture-capital-investments/>

<sup>15</sup> Kts. mm. Vesset (2013), Big Money for Big Data. [https://idc-community.com/groups/it\\_agenda/business-analytics-big-data/big\\_money\\_for\\_big\\_data](https://idc-community.com/groups/it_agenda/business-analytics-big-data/big_money_for_big_data)

<sup>16</sup> Novet (2013), Big data startups pull in big money in 2013. <http://venturebeat.com/2013/12/09/big-data-startups-pull-in-big-money-in-2013/>

hyödyntämisen olevan tulevaisuudessa tärkeä osa vähintään yhden osastonsa toimintaa. Suuryritysten big datan hyödyntäminen tulee tutkimuksen mukaan olemaan huomattavasti yleisempää kuin pk-sektorilla.<sup>17</sup>

**USA:** Alan kärkimaassa on panostettu pelkästään hallinnossa big data -kehitykselle satoja miljoonia dollareita. Big data -alalla toimivilla yrityksillä on selkeä etumatka esimerkiksi eurooppalaisiin yrityksiin verrattuna.

Obaman hallinto julkaisi toukokuussa 2014 big dataa käsittelevän raportin<sup>18</sup>, jossa kohdistetaan erityistä huomiota yksityisyydensuojaan. Raportissa todetaan, että eri maiden erilaiset lainsäädäntökehdikot voivat osaltaan aiheuttaa haasteita big datan hyödyntämiselle. Raportissa arvioidaan, että eurooppalainen tietosuojalainsäädäntö voi kattavuutensa vuoksi muodostaa haasteita innovaatiotoiminnalle, kun USA:ssa sääntely on pyritty kohdistamaan yksityisyydensuojan osalta kriittisempiin osa-alueisiin kuten terveyttä koskeviin tietoihin.

**Yhdistynyt kuningaskunta:** Hallitus on määritellyt asiantuntija-arvioiden pohjalta big datan yhdeksi kahdeksasta maan taloudellisen tulevaisuuden kannalta keskeisestä teknologiasta. Tähän liittyen hallitus julkaisi vuonna 2013 huomattavia investointilupauksia sisältävän toimenpideohjelman "Seizing the data opportunity – A strategy for UK data capability"<sup>19</sup>. Toimenpideohjelma keskittyy pitkälti big dataan liittyvään alan koulutuksen, kuten data-analytiikan, tarpeiden ja puutteiden arvioimiseen ja korjaamisen sekä muun muassa opetusohjelmien uudistamiseen ja erilaisten harjoitteluiden lisäämiseen. Lisäksi ohjelma keskittyy big datan hyödyntämiseen liittyvän infrastruktuurin vahvistamiseen ja yritysten tietoisuuden lisäämiseen.

Lontoosta on kehitymässä suuri big data -alan eurooppalainen hubi amerikkalaisten alan johtavien alueiden rinnalle. Lontoon läheisyydessä (40 km säteellä) työskentelee jo 54 000 henkeä big dataan liittyvien työtehtävien parissa. New Yorkin ja San Franciscon sekä Piilaakson vastaavat luvut ovat 57 000 ja 98 000 henkilöä.<sup>20</sup>

**Saksa ja Ranska:** Big data nähdään teollisuuden kilpailukyvyyn avaintekijänä. Maat ovat huomioineet 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.<sup>21</sup>

**EU:** Datan hyödyntämismahdollisuudet ja big datan potentiaali ovat nousseet vahvasti esille myös EU:n digitaalistrategiaan ja uuteen Horizon 2020 –ohjelmaan liittyvissä hankkeissa ja rahoitusinstrumenteissa. Big data ja avoin data ovat komission kasvua ja uusia työpaikkoja hakevan Digital Agenda for Europe -kärkihankkeita.

<sup>17</sup> IDG Enterprise Research 2014: Key Findings. <http://www.idgenterprise.com/report/big-data-2>; Columbus (2014), 2014: The Year Big Data Adoption Goes Mainstream In The Enterprise. <http://www.forbes.com/sites/louiscolumbus/2014/01/12/2014-the-year-big-data-adoption-goes-mainstream-in-the-enterprise/>

<sup>18</sup> The White House Executive Office of The President (2014). Big Data: Seizing opportunities, preserving values. [http://www.whitehouse.gov/sites/default/files/docs/big\\_data\\_privacy\\_report\\_5.1.14\\_final\\_print.pdf](http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf)

<sup>19</sup> HM Government, Seizing the data opportunity – A strategy for UK data capability (2013). <https://www.gov.uk/government/publications/uk-data-capability-strategy>

<sup>20</sup> Mandel & Liebenau (2014), London: Digital City on the Rise. <http://mikebloomberg.com/files/London-Digital-City-On-The-Rise.PDF>

<sup>21</sup> Työryhmän työtä varten koottu katsaus muiden maiden big data-strategioista. <http://www.lvm.fi/lvm-mahti-portlet/download?did=137476>

Komissio julkaisi 2.7.2014 tiedonannon big data -strategiastaan ”Kohti menestyvää data-vetoista taloutta”<sup>22</sup>. Big data nähdään tiedonannossa datavetoiseen talouteen siirtymisen avaintekijäksi ja keskeiseksi eurooppalaisen kilpailukyvyn lähteeksi. Komission tiedonannossa nostetaan esiin myös monia tulevia toimenpiteitä big datan hyödyntämisen ja – kyvykkyyden edistämiseksi Euroopan tasolla. Tulevat toimenpiteet liittyvät muun muassa laajempaan toimijoiden yhteistyöhön sekä osaamiskeskusten perustamiseen. Komissio tulee myös hakemaan eurooppalaisia kärkialoitteita aloilta, joilla voidaan saavuttaa suurimpia yhteiskunnallisia hyötyjä. Komission toimenpiteet ja päämäärät ovat hyvin linjassa myös tässä raportissa esitettyjen ehdotusten kanssa.

**OECD:** OECD nostaa esiin big data –analyysin mahdollisuuksia erityisesti New Sources of Growth: Knowledge-Based Capital –teemaan liittyvissä projekteissa ja selvityksissä. Näissä on korostettu vahvasti big datan potentiaalia eri alojen innovaatioiden ja kasvun tuojana muun muassa hallinnossa, terveysalalla sekä liikenteessä. Lisäksi OECD kiinnittää huomiota muun muassa osaamispuitteisiin sekä infrastruktuuriin liittyviin haasteisiin. Samalla OECD korostaa myös tarvetta arvioida datan käytön vaikutuksia ja arvoa, jotta uusia käytäntöjä ja tiedonhallintamenetelmiä voidaan kehittää.<sup>23</sup>

## 1.7 Suomen nykytila ja asema kilpailussa

Suomella on hyvät edellytykset vastata big datan haasteeseen: Kansainvälisten vertailututkimusten mukaan Suomi on maailman digitaalisesti edistyneimpien maiden joukossa erityisesti mitä tulee teknisiin valmiuksiin<sup>24</sup>. Suomessa on maailmanlaajuisesti poikkeuksellisen korkeatasoiset digitaaliset perusrekisterit ja aiheeseen liittyvää osaamista, joiden pohjalta voidaan suhteellisen nopeasti rakentaa kilpailukykyistä big data -toimintaa. Esimerkiksi Suomen laadukkaat terveydenhuollon tietovarannot voivat tarjota vastuullisesti käytettynä suuren innovaatiopotentiaalin.

Suurten data-aineistojen hyödyntämisen ja kilpailukyvyn kannalta on löydettävissä monia potentiaalisia alueita. Perinteisesti Suomessa on osaamista mm. lääketieteellisessä tutkimuksessa, mobiiliteknologioissa, peliteollisuudessa ja ympäristön monitoroinnissa, jotka kaikki ovat hyvin dataintensiivisiä ja sen monimuotoiseen analyysiin perustuvia aloja.

Big data –analyyseistä on saatavissa hyötyjä myös niillä yhteiskunnallisesti ja taloudellisesti merkittävillä alueilla, joilla olemme selkeästi kehityksessä jäljessä tai joilla datalähtöisyys ei ole vielä selkeä osa toiminnan johtamista. Erityisesti teollisen internetin odotetaan tarjoavan uusia mahdollisuuksia, ja esimerkiksi Konecranes ja Outotec ovat alan suomalaisia edelläkävijäryhmiä.

Suomella on vahvaa menetelmä- ja it-osaamista, jota muuntamalla ja päivittämällä sekä koulutuksen, tutkimuksen ja asiantuntemuksen jakamisen kautta saataisiin nykyistä paremmin hyödynnetyksi big data -kehitystyössä. Suomella on vahvuuksia myös tieteelli-

<sup>22</sup> Komission tiedonanto Euroopan parlamentille, neuvostolle, Euroopan talous- ja sosiaalikomitealle ja alueiden komitealle. Kohti menestyvää datavetoista taloutta. COM(2014) 442 final. <http://ec.europa.eu/transparency/regdoc/rep/1/2014/FI/1-2014-442-FI-F1-1.Pdf>

<sup>23</sup> OECD: New Sources of Growth. Knowledge-based Capital. <http://www.oecd.org/sti/inno/knowledge-based-capital-synthesis.pdf>

<sup>24</sup> Etila, Digibarometri 2014. <http://www.etla.fi/uutiset/digibarometri-2014-huipputason-edellytyksilla-keskinkertaisia-tuloksia/>; WEF, Global Information Technology Report 2014. [http://www3.weforum.org/docs/WEF\\_GlobalInformationTechnology\\_Report\\_2014.pdf](http://www3.weforum.org/docs/WEF_GlobalInformationTechnology_Report_2014.pdf)

sessä laskennassa ja tiedon käsittelymenetelmiin liittyvässä osaamisessa. Big datan ympärille syntyvien arvoketjujen ja ekosysteemien kannalta eri toimijoiden ja vahvuuksien tunnistaminen on oleellista.

Big datan hyödyntäminen julkisella sektorilla on vasta alkutekijöissään, mutta se tarjoaa suuria mahdollisuuksia niin palvelujen kuin prosessienkin parantamiseen ja tehostamiseen sekä uusien toimintatapojen käyttöönotolle. Suomi on ollut edelläkävijämaita julkisen sektorin tietovarantojen avaamisessa. Avoimuuden ja hyvälaatuisten julkisten tietovarantojen saatavuuden kulttuuria tulisi hyödyntää myös big data -kehitystyössä. Avoimen ja big datan sekä niin julkisten kuin yksityistenkin data-aineistojen yhdistelyssä piilee suuri potentiaali.

Big datassa on huomioitava erityisesti erilaisten datojen vaatimukset datan käytön ja hallinnan suhteen. Esimerkiksi terveysdatan käsittely henkilötietona tarjoaa hyvin erilaisia käyttömahdollisuuksia kuin esimerkiksi laitteista saatava anturidata.

Suomi on korkean tietosuojaan maa. Nykyisessä maailmassa tätä voidaan pitää merkittävänä vahvuutena ja kilpailutekijänä, jota ei saa vaarantaa. Suomen korkealaatuinen langaton ja kiinteä viestintäinfrastruktuuri antaa hyvän pohjan erilaiselle teolliseen internetiin ja esineiden internetiin (Internet-of-Things, IoT) liittyvälle kehitykselle. Viestintäinfrastruktuurin riittävyys tulevaisuuden tarpeisiin on varmistettava.

Hitaalla reagoinnilla saatetaan edellä mainitut vahvuudet hukata. Niin hallinnossa, tutkimuksessa kuin yrityksissäkin täytyy tehdä tarvittavia muutoksia, jotta uusia ratkaisuja voidaan kehittää ja ottaa käyttöön. Maailmalla big dataan panostetaan paikoin jo vahvasti. Vaarana on yhä dataintensiivisemmäksi muuttuvan kehityksen keltasta putoaminen.

## **2. Edellytykset**

Tässä luvussa kuvataan tärkeiksi tunnistettuja, big datan hyödyntämisen yleistymisen kannalta oleellisia yleisiä edellytyksiä. Luvussa mainittujen edellytysten tulee olla kunnossa, jotta big data -ratkaisuja voidaan soveltaa eri aloille ja saavuttaa tavoiteltuja hyötyjä.

### **2.1 Datatietoisuus**

Datatietoisuudessa on kyse datan potentiaalisen arvon ja toisaalta riskien ymmärtämisestä. Hyöty ei aina ilmene heti ja ole saatavissa suoraviivaisesti, mutta tulevaisuudessa omaa dataa esimerkiksi muiden reaaliaikaisten tietolähteiden kanssa yhdistelemällä saatetaan saavuttaa lisäarvoa. Analytiikalla datasta on mahdollisuus saada merkittäviäkin oivalluksia. On osattava kysyä oikeita kysymyksiä ja hakea oikeita asioita. Datatietoisuus koskee niin yrityksiä, tutkimusta, hallintoa kuin yksilöitäkin.

Nykyaikaisen analytiikan ja big datan omaksuminen osaksi jokapäiväistä työtä on keskeinen askel kohti tiedolla johtamisen yrityskulttuuria. Toimintaympäristön ja uuden teknologian jatkuva kehitys, esimerkiksi tallennuskapasiteetin halventuminen, vaatii omien kyvykkyyksien ja valmiuksien säännöllistä uudelleentarkastelua. Tiedon keruun ja ana-

lysoinnin tulee olla mahdollisimman luonteva osa organisaatioiden muuta toimintaa, ei erillinen saareke.

Datamassoista voidaan joka vuosi etsiä 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.

Perinteisesti yritysorganisaatiot ovat tottuneet tarkastelemaan oman liiketoimintansa tuottamaa dataa ja siitäkin erityisesti esivalittua osajoukkoa, liiketoiminnan ydintietoa. Tämä tieto on kerättyä erilaisissa toiminnanohjaus- ja muissa tietojärjestelmissä, joissa sekä datan tuntemus, että sen analysointiin käytettyjen välineiden osaaminen on vahvaa.

*”Big datan hyödyntäminen ei niinkään tarkoita suurta datamäärää vaan suuria ja uusia hyödyntämismahdollisuuksia”*

Nyt uusia datalähteitä täytyy osata hyödyntää laajemmin, huomioiden sekä ulkoiset että sisäiset lähteet.

Ydintoimintaa 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 antureita?

Näihin mahdollisuuksiin tarttuminen kartuttaa oman toiminnan tuottamaa datavarantoa entisestään ja saattaa nostaa big datan teknologioiden omaksumisen varsin ajankohtaiseksi.

Datatietoisen organisaation ei tarvitse rajoittaa tiedonjanoaan ainoastaan omien järjestelmien keräämään dataan, vaan se voi katsoa ympäröivää ekosysteemiä laajemmin. Datan mahdollisuuksia on syytä arvioida kumppaniverkoston kanssa, tunnistuen synergioita, datan vaihdon mahdollisuuksia tai jopa uuden liiketoiminnan perustamisen potentiaalia datavarantojen yhdistämisen myötä. Myös tutkimuksessa on oleellista ymmärtää uusien datalähteiden, teknologioiden ja menetelmien tuomat hyödyntämismahdollisuudet.

Julkisella sektorilla big datan mahdollisuuksien ja mahdollisten riskien tunnistaminen on oleellista. Eri hallinnonalojen tavoitteissa ja erilaisissa toimintastrategioissa on monia asioita, joissa voitaisiin big dataan ja tiedon luovaan yhdistelemiseen perustuvien ratkaisuiden avulla saavuttaa muun muassa merkittäviä säästöjä, tehdä informoituneempia päätöksiä, ja selkeästi parantaa kansalaisille suunnattuja palveluita.

*”Miten big data vaikuttaa meihin tiedon ja ymmärryksen tasolla?”*

Eri puolilla hallintoa tulisikin tunnistaa, kuinka datalähtöisten ratkaisujen avulla voidaan etsiä helpotusta myös suuriin yhteiskunnallisiin ongelmiin, kuten ikääntymiseen sekä harmaaseen talouteen. Ylipäätään erilaisissa hallinnon ja yritysten toimintasuunnitelmissa ja –strategioissa erilaisten tietoaaineistojen hyödyntämisen mahdollisuudet tavoitteiden toteuttamiseksi tulisi huomioida laajasti. Tähän tarvitaan myös ymmärrystä siitä, mitä mahdollisuuksia tiedon käyttö voi tarjota, niin organisaatiolle itselleen, muulle hallinnolle kuin koko yhteiskunnalle laajemmin.

Yksilöiden datatietoisuudella tarkoitetaan muun muassa ymmärrystä omien tietojen jakamiseen liittyvistä riskeistä. Vaikka yksittäisestä jaetusta tiedosta ei vielä voikaan päätellä mitään, voi analyysimenetelmiä käyttämällä ja erilaisia tietoaaineistoja yhdistelemällä saada hyvin tarkasti selville mm. sen, mitä/kuka on kyseessä, miten joku on toiminut ja jopa miten hän tulee seuraavaksi toimimaan. Tällaiset tiedot ovat kullannarvoisia monessa toiminnassa. Yksilön ymmärrys muun muassa oman sijaintitietonsa jakamisesta, jo kerätyistä tiedosta ja profiloinnin erilaisista mahdollisuuksista ja riskeistä olisi hyvä olla olemassa. Lisäksi ymmärrys datan luotettavuudesta korostuu.

**Sosiaali- ja terveysministeriön Sote-tieto hyötykäyttöön -strategiatyössä big data on ymmärretty merkittävimäksi keinoksi tavoitteiden saavuttamisessa:**

"Big data -ratkaisuja kehitetään tukemaan suurten tietomassojen hyödyntämistä, edistynyttä erityyppisten tietojen yhdistelyä, kliinisen potilas- ja genomitiedon sekä kuvantamistietojen analysointia sekä tietojen yhdistelyä kansallisten ja henkilökohtaisten tietovarantojen kanssa. Arvioidaan kansallisen big data alustan toteuttaminen tutkimuskäytön edistämiseksi (esim. toteutus kansallisessa osaamiskeskuksessa, johon otettaisiin mukaan eri yritysten big data teknologioita ja analytiikka-työkaluja)."

## 2.2 Koulutus ja osaaminen

Osaaminen on big datan hyödyntämisen kannalta avainasemassa. Jo nyt on nähtävissä osaamisen kapeikkoja. Muun muassa erilaisten analyysimenetelmien osaajista on tulevaisuudessa kasvavaa pulaa. Pulaa on myös big datan käsittelyä osaavista ohjelmoijista. Samoin tarvitaan yritys- ja julkishallinnon johtajia, jotka ymmärtävät data-analyysin mahdollisuudet päätöksenteon tukena, tuotekehitystyössä sekä kilpailukyvyyn turvaajana. Johdon tueksi tarvitaan osaavia, koulutettuja henkilöitä, jotka osaavat muuntaa analyysin tulokset johtajien ymmärtämään muotoon.

Koulutuspuutteet voivat muodostua merkittäväksi esteeksi big datan laajemmalle hyödyntämiselle. Big datan murrokseen on kyettävä vastaamaan nopeilla ja nykyistä kohdentummilla koulutustoimilla. Big data edustaa uutta hyppäystä datatutkimuksen ja -soveltamisen alalla, jolla perinteiset koulutusohjelmat eivät vastaa enää ammatillisiin tarpeisiin. Eri toimialojen omat sovellukset edellyttävät soveltavaa osaamista myös muista aloista (esim. kaupan, teollisuuden tai oppimiseen liittyvä osaaminen).

*"90 % osaamiskysymys"*

Big datan hyödyntäminen mahdollistaa ja edellyttää myös työprosessien uudistamista. Työelämässä toimivista tietualan ammattilaisista suurella osalla ei ole sellaisia ammatillisia valmiuksia, joita big datan hyödyntäminen liiketoiminnaksi edellyttää. Työelämässä vaadittavat taidot muuttuvat digitalisaation myötä. Datataidot tulisikin ymmärtää osaksi elinikäistä oppimista.

**Perus- ja toisen asteen koulutus:** Koulutuspuutteisiin täytyy vastata jokaisella koulutustasolla. Tämän tulee tapahtua jo lähtövalmiuksia tarjoavasta peruskoulusta alkaen. Perustana olevia matemaattisia, tilastotieteellisiä sekä ohjelmointitaitoja tulisi painottaa opetussuunnitelmissa jo hyvin varhain, jotta tulevat osaamistarpeet tulisivat paremmin katetuksi pitkällä tähtäimellä. Tärkeiden tietotyökalujen opetusta tulee esimerkiksi keskiasteella lisätä ja luoda pohjaa mm. tiettyjen tietotyökalujen hallinnassa. Lisäksi kaikessa koulutuksessa tulisi korostaa datalähtöisempää ja empiiristä, kokeilevaa lähestymistapaa.

*”Peruskoulun tulee antaa taidot datankäsittelyyn”*

Opettajien koulutus on tärkeässä asemassa, jotta jo peruskoulussa voidaan harjoitella tietojen hyödyntämistä. Opettajien osaamisen rajoitteet ja riittävyys voivat muodostaa big data -osaamisen kehitykselle haasteen kaikilla koulutustasoilla. Yhtenä ratkaisuna voivat olla erilaiset uudet verkko-oppimistyökalut sekä massaverkkokurssit (MOOC, massive open online course). Näiden hyödyntäminen vaatii esimerkiksi yliopistoilta halua yhteistyöhön. Toisaalta dataan perustuvilla koulutusratkaisuilla ja –järjestelmillä sekä sovelluksilla on myös hyvä vientipotentiaali.

**Korkeakoulutus:** Suomessa on käynnistetty tai 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. Liitteessä 4 on esitelty Suomessa jo järjestettävää ja alkamassa olevaa big dataan liittyvää korkeakouluopetusta. Lausuntokierroksella kerättyä tietoa hyödyntää jatkossa opetus- ja kulttuuriministeriö.

*”Kilpailu digitaalisesta osaamisesta on kovaa - neljäsosa yrityksistä häviää osaamisvajeen vuoksi”*

Kaikkien suomalaisten korkeakoulujen, jotka eivät vielä ole havahtuneet big datan vaikutuksiin, tulisi huomioida kehitys ja pikaisesti saada datalähtöisyys ja data-analyysi näkymään eri alojen koulutuksissa. Näissä koulutusohjelmissa on välttämätöntä tehdä yritysten kanssa mahdollisimman ennakoivaa ja tiivistä yhteistyötä, jotta tulevat työelämän tarpeet saadaan edes osittain ennakoivasti katetuiksi myös suoraan käytännön taidoilla varustetuilla ammattilaisilla. Alan toimijoiden, yritysten ja koulutusorganisaatioiden tulee pystyä hyvässä yhteistyössä määrittelemään laajojen tietoaineistojen hyödyntämisen vaatimia koulutustarpeita. Laajasti katsoen koulutuksen täytyy pystyä vastaamaan määritettyihin osaamistarpeisiin.

### Työelämälähtöinen osaamistarve-esimerkki: E-Skills UK

E-Skills UK:n tekemässä selvityksessä big datan edellyttämistä taidoista saa hyvän yleiskuvan niistä teknisistä taidoista, joita big datan käyttäminen osana liiketoimintaa edellyttää. Vaikka big datan määritelmää ei olekaan hakattu kiveen, on kysymys kuitenkin hyvin läheisesti ICT-teknologiaan liittyvistä työtehtävistä, jotka edellyttävät asianmukaisia taitoja. Kiistatonta on, että tieto, data, lisääntyy ja monipuolistuu eksponentiaalisesti, ja tämän tiedon käsittely, analyysi ja johtaminen vaativat spesifejä teknisiä taitoja. Raportissa arvioitiin vaadittavia big dataan liittyviä taitoja työpaikkailmoitusten ja niissä käytettyjen termien avulla. Samalla pyrittiin arvioimaan myös tulevaisuuden tarpeita big data -osaamistarpeita Isossa-Britanniassa vuoteen 2017 asti.

Raportti jakaa big data -tehtävät karkeasti roolien mukaan neljään eri ryhmään: Ohjelmistokehittäjät (Developers), tietoarkkitehdit (Architects), analyytikot (Analysts) ja johtajat (Administrators). IT-osaamisen erikoisalueista erityisesti tarvittiin NoSQL-, Oracle-, Java- ja SQL-osaamista. Myös ohjelmistokehittämisen metodologiselta puolelta esitettiin vaatimuksia, mm. ketterän ohjelmistokehityksen taitoja, testivetoisen kehittämisen kykyjä, big datan käyttöönottoon ja muuntamiseen keskittyviä ETL-taitoja (Extract, Transform and Load) sekä CSS-taitoja.

Ohjelmistokehittäjien tehtävänimikkeellä esiintyneisiin tehtäviin (Business Intelligence Developer, Web Developer, Software Developer, Business Developer, Analyst Developer, Applications Developer, Database Developer ja Front-End Developer) teknisissä taitovaatimuksissa esiintyivät yleisesti TDD-, CSS- ja ketterän ohjelmistokehityksen taidot. Näiden lisäksi tehtävissä vaadittiin spesifejä taitoja NoSQL:stä (erityisesti MongoDB), Java-osaamista, SQL-, JavaScript-, MySQL-, Linux-, Oracle-, Hadoop- (especialy Cassandra), HTML- ja Spring-osaamista.

Data-arkkitehtien eri tehtävänimikkeet (mm. Solutions Architect, Data Architect ja Business Intelligence Architect) pitivät työpaikkailmoitusten mukaan sisällään useita IT-teknisiä erityistaitoja, kuten Oracle (erityisesti BI EE), Java, SQL, Hadoop ja SQL Server. Yleisellä tasolla data-arkkitehteiltä edellytettiin datamallinnus-, ETL-, erityis-arkkitehtuuri-, avoimen lähdekoodin ja data-analyysin taitoja. Varsinaisille data-analysioijille yhteisiä taitoja olivat eri tehtävänimikkeidensä (Business Analysts, Data Analysts, Business Intelligence Analysts ja Support Analysts) mukaisesti datamallinnus-, ETL, analyttiset ja tieteanalysitaidot.

Keskimäärin informaatio- ja telekommunikaatiotehtävissä olevien määrän on arvioitu kasvavan Isossa-Britanniassa 2,5 % vuosittain vuoteen 2017 asti. Big data-osaajien työllisyys kuitenkin kasva huomattavasti reippaammin raportin ennusteen mukaan. Raportissa löydetään huomattava kasvu NoSQL-osaamisen tarpeessa raporttia edeltävän kahden vuoden ajalta: kasvun todettiin olevan 1600 %. Samoin Hadoop-osaajien tarve oli kasvanut kahdessa vuodessa 700 %. Hadoopissa käytettävän NoSQL-tietokantakomponentin (Hbase) osaamistarve oli kasvanut samana aikana 2370 %. Kokonaistyöpaikkojen määrän arvioitiin kasvavan tutkimusta seuranneella ajanjaksolla 2012-2017 alimman skenaarion mukaan yhteensä 65 % (13 % vuosittain) ja korkean kasvun skenaarion mukaan jopa 117 % viidessä vuodessa (23 % vuosittain).

Lähde: [https://www.e-skills.com/Documents/Research/General/BigDataAnalytics\\_Report\\_Jan2013.pdf](https://www.e-skills.com/Documents/Research/General/BigDataAnalytics_Report_Jan2013.pdf)

Suurten tietoaisteistojen hyödyntämisen alueella tulisi lisätä monialaista koulutusta. Datalta täytyy osata kysyä oikeita kysymyksiä, joten erilaisten taitojen ja erilaisen osaami-



sen yhdistäminen on välttämätöntä. Niinpä data-analytiikkaa ja datan "ymmärtämistä" tulisi sisältyä melkein alan kuin alan koulutukseen, aina kauppatieteellisestä yhteiskunta- ja sosiaalitieteisiin. Pelkkä tekninen osaaminen ei riitä, vaan tarvitaan ymmärrystä substanssialoista sekä kykyä yhdistellä asioita ja aloja luovasti.

Monella alalla käytettävissä olevan tiedon hyödyntäminen toisi huomattavaa lisäarvoa. Samalla on tärkeää kouluttaa osaajia myös tiedon hyödyntämisen, yksityisyyden ja tietosuojan alueilla. Pelkästään teknologiaan liittyvä koulutus ei riitä, vaan tarvitaan osaamista ja tietoa siitä, mitä dataa saa kerätä ja käyttää. Myös informaatio-oikeudellisen osaamisen tarve kasvaa, kun erilaista tietoa kerätään ja hyödynnetään yhä enemmän. Esimerkiksi EU:ssa valmisteilla oleva tietosuojalainsäädäntö saattaa tuoda yrityksille merkittäviä sanktioita henkilötietojen virheellisestä käytöstä<sup>25</sup>. Nykyinenkin lainsäädäntö antaa yrityksille mahdollisuuden käyttää henkilötietoa sisältävää dataa vain siihen tarkoitukseen, johon se on alunperin kerätty, mikä luonnollisesti rajoittaa datan käyttöä huomattavasti.

Erilaisilla harjoittelu- ja tutustumismahdollisuuksilla on oma tärkeä osansa. Työelämän ja julkishallinnon tulisi tarjota data-analytiikkaan koulutettaville opiskelun aikana mahdollisuuksia erilaisiin harjoittelujaksoihin ja käytännön harjoitustöihin. Tämä koskee niin suoraan käytännönläheiseen työhön, kuten koodaamisen sekä vaikkapa laajemmin data-analytiikkaan ja data management -aihepiiriin keskittyviä opiskelijoita. Myös yritysten aineistoja tulisi saada koulutuskäyttöön. Tällöin opiskelijat voisivat oppia empiirisemmin suoraan käytännön ongelmien parissa.

**Täydennyskoulutus:** Yritykset, julkinen sektori ja muut toimijat tarvitsevat välittömästi henkilöstön osaamista uudistavaa täydennyskoulutusta ja kursseja big datan hyödyntämiseksi. Uusilla täydennyskoulutusohjelmilla ja -malleilla voidaan tukea tietoteknologia-alalla osaamisen uudistamista ja saumattomia ammatin vaihdoksia.

Ongelmana on jälleen osaavien kouluttajien pieni määrä. Erilaisten massaverkkokurssien (MOOC) ja muiden verkkotyökalujen käyttö voi helpottaa akuuttia pulaa. Big data -kehityksen laaja-alainen hyödyntäminen vaatii jatkuvaa osaamisen päivittämistä niin henkilöstöltä itseltään kuin johdolta. Akuutti osaajatarve täytyy pitkälti ratkaista yritysten omalla henkilöstökoulutuksella ja tukemalla osaajien keskinäistä verkostoitumista välittömän, jo olemassa olevan osaamisen ja tietoisuuden leviämiseksi.

Uusia työnkuvia muodostuu jatkuvasti. Kaikkea big datan hyödyntämiseen osaamista ei välttämättä voida kouluttaa samalle henkilölle. Erilaisia vahvuuksia sisältävät monialaiset tiimit tulevat työs-

*"Big data osaamista voisi kehittää joustavasti työn ohessa ja eri alojen osaajat voisivat täydentää toisiaan"*

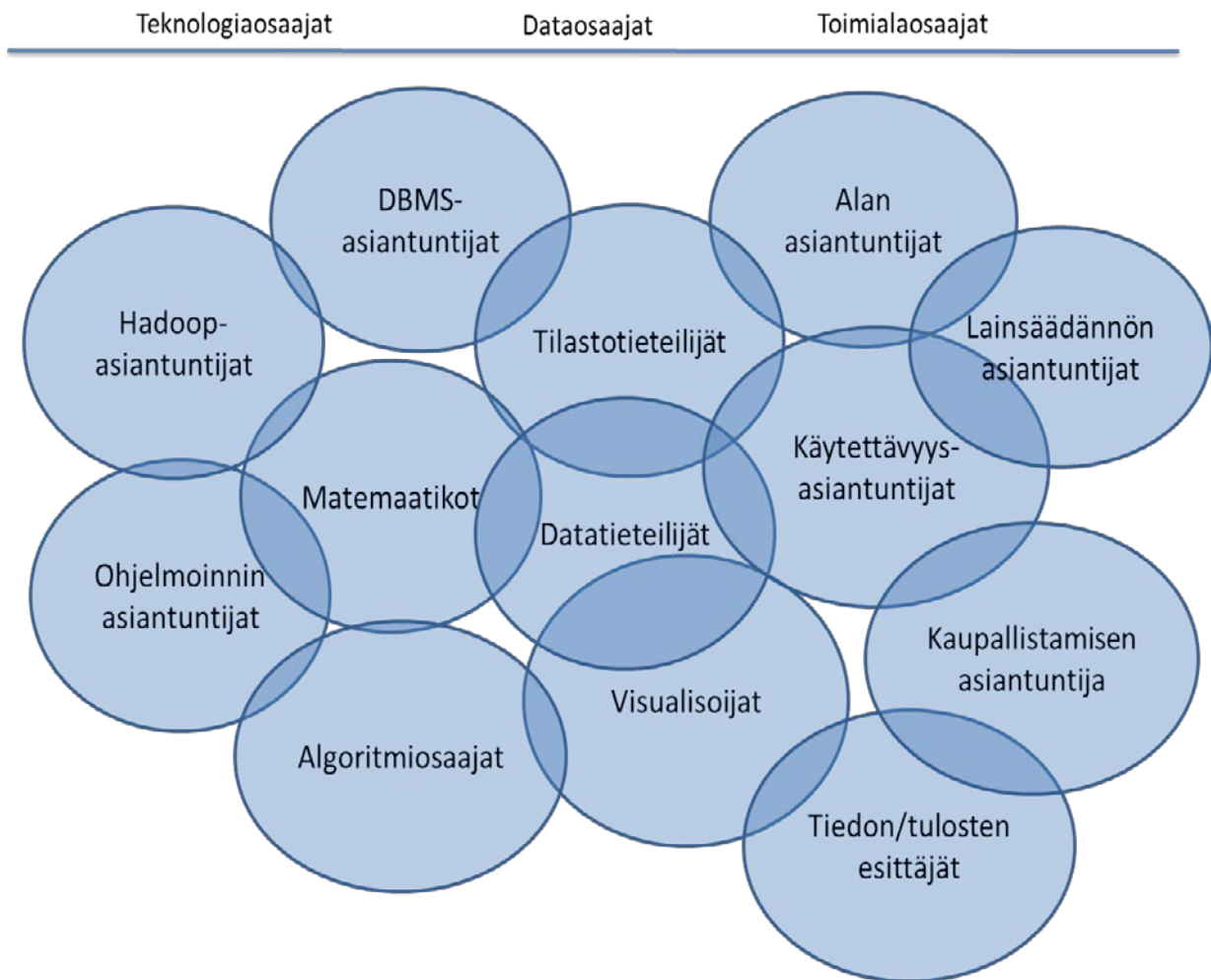
kentelemään big datan parissa. Niinpä jo koulutuksessa erilaista tiimityöskentelyä täytyy tukea ja erilaisen koulutustaustan omaavia saattaa yhteisten ongelmien eteen. Myös yrittäjyyteen ja esimerkiksi käytettävyyssuunnitteluun sekä palvelumuotoiluun liittyvää kou-

<sup>25</sup> Ehdotus Euroopan parlamentin ja neuvoston asetukseksi yksilöiden suojelusta henkilötietojen käsittelyssä sekä näiden tietojen vapaasta liikkuvuudesta (yleinen tietosuoja-asetus) COM/2012/011. <http://eur-lex.europa.eu/legal-content/FI/TXT/?qid=1403675298403&uri=CELEX:52012PC0011>

lutusta tarvitaan samaan yhteyteen, jotta ratkaisujen kaupallistaminen onnistuisi nopeammin.

Oheisessa kuvassa 4 on esimerkki siitä, kuinka laajaa osaamista big datan hyödyntäminen vaatii ja kuvaus big data -tiimin erilaisista rooleista. Lähtökohtaisesti erilaisista rooleista huolimatta big dataa tehokkaasti hyödyntävissä tiimeissä myös yksilöiden moniosaajuus ja roolien sekoittuminen korostuvat. Datatieteilijöiden ja data-asiantuntijoiden tulee oman ydinosaamisensa olla lisäksi usein taitavia ohjelmoijia ja omata vahva substanssiymmärrys ongelmanratkaisun kohteena olevasta toimialasta.

**Kuva 4: Esimerkki big data -tiimin osaamisesta**



Yhteistyö oppilaitosten kanssa ja muun muassa käytännön harjoittelut auttavat tulevaisuudessa osaajapulaan. Akuutti (seuraavat pari vuotta) osaajatarve täytyy kuitenkin ratkaista pääosin muilla keinoilla.

## 2.3 Tutkimus

Big data -menetelmien hyödyntäminen tutkimustoiminnassa on osin välttämätöntä, tutkimustiedon ja potentiaalisen tutkimusdatan määrän ja muotojen kasvaessa rajusti. Big data -menetelmien käyttö on suomalaisessa tutkimuksessa vielä suhteellisen vähäistä. Nykyiset big data -teknologiat ovat suurilta osin kaupalliselta puolelta. Tutkimuksessa tarvitaan myös erilaisia, nimenomaan tutkimuskäyttöön räätälöityjä teknologioita. Samalla tarvitaan menetelmien siirtoa tutkimuksen ja yrityssektorin välillä.

Big datan perustutkimus ja menetelmiä soveltava tutkimus käyvät käsi kädessä. Käytännön ongelmien ratkaiseminen ruokkii big dataa hyödyntävien tutkimuksellisten ratkaisuiden kehittämistä. Samalla saadaan ajoittain myös yleisemmin sovellettavia menetelmiä, joita käytetään uusien ja erilaisten big data -ongelmien ratkaisemiseen. Näin big dataan liittyvät perustutkimus ja soveltava tutkimus ruokkivat toisiaan.

*”Tutkimusparadigma ei ole muuttunut, mutta tutkimuksessa nyt erilaiset syntyprosessit”*

Perinteisesti Suomessa on huippuluokan osaamista esimerkiksi tieteellisestä laskennasta, algoritmikasta, 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.

Big dataa voidaan hyödyntää monilla tutkimusaloilla. Esimerkiksi biotieteiden, tähtitieteen, fysiikan ja vaikkapa psykologian (psykonometria) ja taloustieteen (ekonometria) aloilla big data -kehityksestä ja -menetelmistä on saatavissa selkeitä hyötyjä ja mahdollisuuksia tutkimuksen kehittämiseen.

Big datan hyödyntämiseen eri tutkimusalueilla liittyy myös paljon kysymyksiä, joiden ymmärtämiseen tarvitaan monialaista tutkimusta. Esimerkkinä tällaisesta monialaisesta aloitteesta on Tampereen teknillisen yliopiston valmisteilla oleva big datan tutkimuskeskus, joka yhdistää kolmen eri TTY:n laitoksen osaamista<sup>26</sup>. Tietoaineistojen hyödyntämiseen liittyvää tutkimusta tehdään yliopistojen lisäksi myös ammattikorkeakouluissa, erilaisissa tutkimuslaitoksissa sekä yrityksissä ja niiden erilaisissa hankkeissa. Kansallisella tasolla tutkimusverkostojen koordinointi ja integrointi antaisi mahdollisuuksia monialaisten ja monipuolisten innovaatioiden syntymiseksi. Yksi tähän liittyvä mahdollisuus on myös suomalaisten tutkimusverkostojen linkittäminen Euroopan- ja globaalin tason tutkimusyhteistyöhön.

Big dataa voidaan käyttää tutkimuksessa monilla tavoilla. Samalla tutkijoiden ja tutkimusryhmien tarpeet datalle 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 tekniseen puoleen. Esimerkiksi tilastotieteilijätaustaiset data-analyytikot taas keskittyvät nimenomaan analyysimenetelmiin. Big datan tutkimukselliset soveltajat, esimerkiksi biologit, ovat nimenomaan oman soveltamisalansa asiantuntijoita. Eri ryhmien

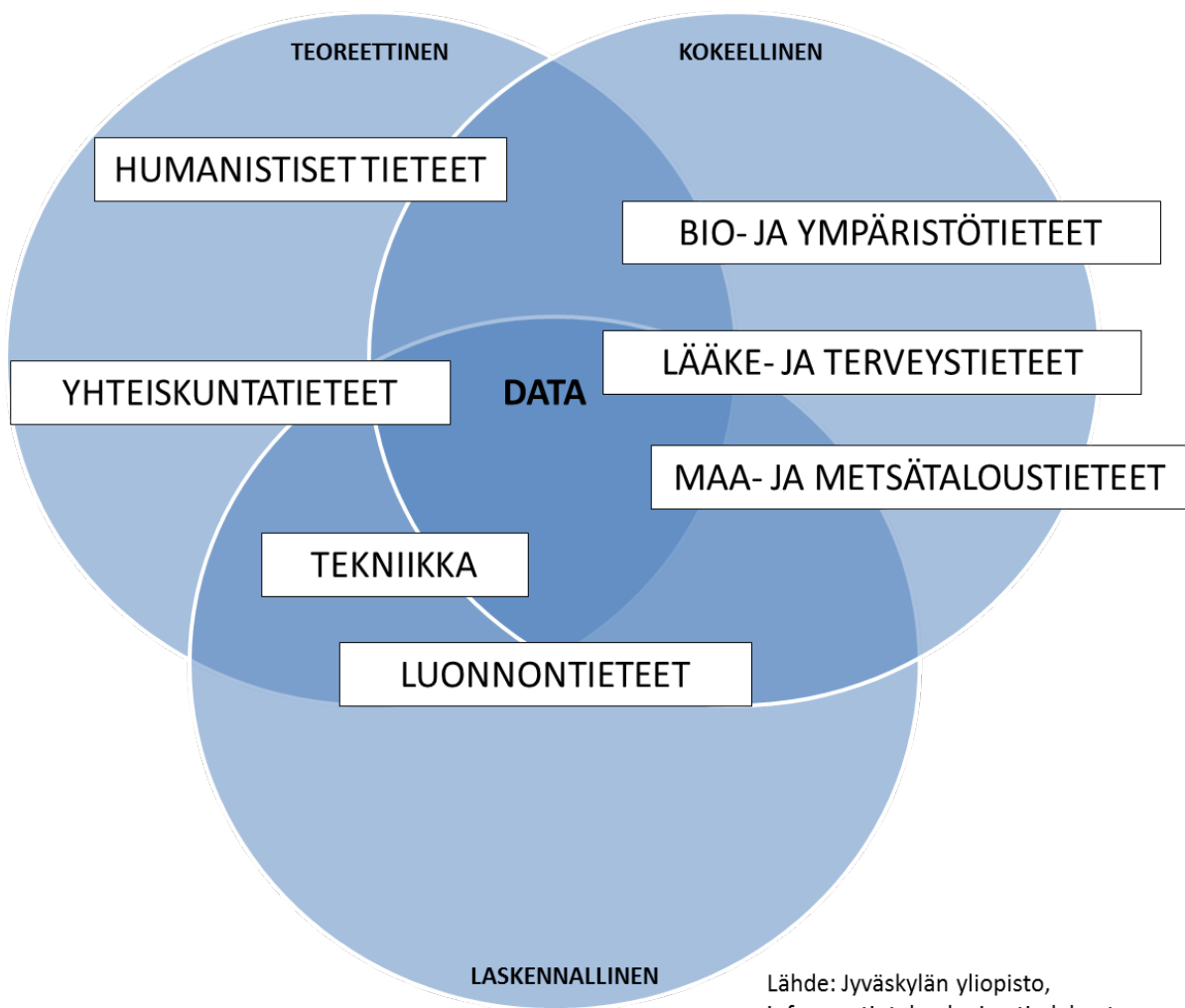
<sup>26</sup> Koskenlaakso (2014), Ei päivää ilman big dataa. <http://www.tut.fi/raipinta/artikkelit/2014/2/ei-paivaa-ilman-big-dataa>

kesken tarvitaan yhteistyötä, jotta uusien big data -menetelmien hyödyt saadaan tutkimuksessa ulosmitattua.

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 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 kehittää yhteistoiminnan muotoja, käytäntöjä ja sovelluksia. Näin kompetenssi rakentaa uutta tiedontuotannon maailmaa kasvaa merkittävästi.

Tutkimukseen käytettävien aineistojen analyysimenetelmien tulee kehittyä sellaisilla tavoilla, että ne hyödyttävät digitaalisia jälkiä jättäviä ihmisiä ja yhteiskuntaa kaupallista tai taloudellista hyötyä laajemmassa mielessä. Aineistoanalyysi voi aidosti tukea kansalaisia esimerkiksi heidän hyvinvointipyrkimyksissään tai arvoprojekteissaan.

**Kuva 5. Big data neljän tiedeparadigman leikkauspisteessä**



Tutkimusnäkökulmasta big dataa voi ajatella myös yhteistoiminnan areenana. Tämä 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. Yhteistoimintaa tarvitaan myös yrityssektorin ja julkisen sektorin kesken esimerkiksi yhteisten tutkimusohjelmien muodossa.

Big data haastaa tutkijoita sekä käyttämään mielikuvitustaan että tekemään eettisesti kestäviä valintoja tutkimustyössään. Yhdysvaltalaiset bioinformatiikan tutkijat ovat jo valmistelleet eettiset säännötöt twitteristä avoimesti saatavan datan käytölle tutkimuksessa<sup>27</sup>. Sääntöihin sisältyvät muun muassa tutkimuksen tavoitteiden, menetelmien ja datankäsittelykäytäntöjen läpinäkyvyys ja saatavuus, twiittaajien anonymiteetin suojaaminen, yksityisten ja poistettujen twiittien käyttämättä jättäminen, ja twiittien kontekstin kunnioittaminen.

**Useat maat** pyrkivät big data -tutkimuksen ykköskaartiin. Maailmalla on perustettu muun muassa big dataan keskittyneitä tutkimuskeskuksia. Iso-Britannia investoi vuonna 2014 Alan Turingin nimeä kantavan big dataan ja algoritmeihin keskittyvän instituutin perustamiseen

Lähteet: <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 tehdään muun muassa Suomen Akatemian huippuyksiköissä ja tutkimusohjelmissa. ALGODAN keskittyi vuosina 2008-2013 algoritmiseen data-analyysiin, COIN data-analyysimenetelmiin. LASTU taas yhdistää menetelmien kehittämisen laskennallisesti vaativien ja tieteellisesti erittäin merkittävien kysymysten tutkimiseen.

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

## 2.4 Infrastrukturi

Verkkoon kytkeytyneiden laitteiden ja niiden lähettämän sekä vastaanottaman datan määrä kasvaa kiihtyvää vauhtia. Cison ennusteiden mukaan vuoteen 2017 mennessä yhteensä noin 19 miljardia laitetta on kytkettynä internetiin<sup>28</sup>. Suurin osa dataliikenteestä tulee tällöin muodostumaan koneiden keskinäisestä (machine-to-machine, M2M) viestinnästä.

Koneiden väliseen viestintään liittyvät verkko-osoitteiden riittävyysongelmat on pitkälti ratkaistu IPv6-protokollan käyttöönoton myötä. Kaiken internetissä (Internet-of-Everything) myös tulevaisuuden verkoilta, niin kiinteiltä kuin langattomilta, vaaditaan

<sup>27</sup> ScienceDaily (2014), Researchers propose tactics for ethical use of Twitter data. <http://www.sciencedaily.com/releases/2014/06/140602115646.htm>

<sup>28</sup> Cisco (2012), Cisco's VNI Forecast Projects the Internet Will Be Four Times as Large in Four Years. <http://newsroom.cisco.com/press-release-content?articleId=888280>

yhä enemmän<sup>29</sup>. Big datan analyysi ja erilaiset ratkaisut perustuvat usein reaaliaikaisen tiedon käyttöön. Pienetkin katkot ja viiveet ovat kriittisiä. Big datan käytön kannalta verkon toimintavarmuus ja erilaisten verkkojen saatavuus on äärimmäisen tärkeitä.

Langattomien verkkojen osalta tulee ennakoivasti allokoida 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 anturitiedolle tarkoitettut "Internet-of-Things"-verkot. San Franciscoon ja Piilaaksoon on jo tulossa tällaiseen tarkoitukseen allokoitu mobiiliverkko<sup>30</sup>.

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

Suurten datamäärien varastointi edellyttää, että datakeskusten tulee toimia mahdollisimman tehokkaasti. Suomen kannalta mahdollisuutena on varmistaa uuden dataintensiivisen liiketoiminnan sijoittuminen maahamme kehittämällä edelleen jo nyt huippuluokkaa olevaa verkkoinfrastruktuuriamme. Tähän kuuluvat muun muassa Itämeren tietoliikennekaapeli sekä Jäämeren tietoliikennekaapeli.

Tietojärjestelmien ja tietojohdamisen kannalta vaatimuksia data-infrastruktuureille samoin kuin ylipäättänsä tietohallinnon kehittämiseksi voidaan johtaa myös useiden tietojärjestelmien haasteellisesta nykytilasta, jossa analysoitava data on tyypillisesti hajanaista ja monimuotoista. Tästä seuraa tarve integroida yhteen useita järjestelmiä ja niiden käyttöön tarvittavia ohjelmistoja.

Tietojärjestelmien ollessa tyypillisesti suunniteltuja vakioitujen ja tiukkaan vartioitujen IT-alustojen varaan, on uusien ohjelmistojen ja liityntöjen asentaminen niihin usein hidasta, mahdollisia tietoturvariskejä unohtamatta.

Datan analysointiin tarkoitettujen järjestelmien tulisi olla lähtökohtaisesti nykyistä modulaarisempia. Toisin sanoen organisaatioiden tulisi mahdollistaa helposti eristettävien ohjelmistokokonaisuuksien toteuttamisen muusta järjestelmästä riippumatta. Keskeistä on kyetä tarjoamaan kuhunkin datan hyödyntämistarpeeseen joustavasti mukautettava kehitysympäristö, jossa käytettävät ohjelmistot voidaan valita tarpeen, eikä alustan vaatimusten mukaisesti.

Yleinen tietoturva ja -hallinta keskittyisivät kehitysympäristöjen väliseen toimintaan ja liikenteeseen. Hallinnollinen valmius modulaaristen ratkaisujen käyttämiseen parantaisi myös muun muassa datanjalostajien kilpailuttamista, koska alustavaatimusten korostamisen sijaan voitaisiin keskittyä alustariippumattomien toteutusten kilpailuttamiseen.

---

<sup>29</sup> WEF, The Global Information Technology Report 2014: Rewards and Risks of Big Data. [http://www3.weforum.org/docs/WEF\\_GlobalInformationTechnology\\_Report\\_2014.pdf](http://www3.weforum.org/docs/WEF_GlobalInformationTechnology_Report_2014.pdf)

<sup>30</sup> Simonite (2014), Silicon Valley to Get a Cellular Network, Just for Things. <http://www.technologyreview.com/news/527376/silicon-valley-to-get-a-cellular-network-just-for-things/>

## 2.5 Tekniset käytännöt ja standardit

Eri organisaatioihin kertyvä data täytyy pystyä ottamaan tehokkaasti käyttöön siellä missä sitä tarvitaan. Tätä tarkoitusta varten tarvitaan yhteisiä sopimuksia ja standardeja siitä, kuinka dataa tallennetaan, jaellaan ja prosessoidaan. Big datan täysimääräisessä hyödyntämisessä oleellista on se, miten erilaisista tietokannoista ja dataseteistä voidaan hakea tietoa, malleja ja tuloksia. Yhteentoimivuus on oleellista.

Tällä hetkellä teknologioiden yhteentoimivuutta edistäviä yhteisiä, maailmanlaajuisessa käytössä olevia, esimerkiksi datan varastointiin ja tiedonsiirtämiseen liittyviä big data -standardeja ei kaikilla osa-alueilla ole. Standardisointia kehitetään eri puolilla maailmaa, asiaa koskeva selvitystyö on käynnissä muun muassa Yhdysvaltain National Institute of Standards and Technologyssa<sup>31</sup>. Tämän lisäksi useissa muissa standardoimisorganisaatioissa pohditaan ja etsitään mahdollisuutta eri teknologioiden yhteentoimivuuden varmistaville big data -standardeille. Yhteentoimivuuden ohella standardisointia selvitetään mm. tietohallinnan ja pilvipalvelujen tietoturvan osalta. Standardointia tarvitaan myös yritysten keskinäisen datan vaihdon mahdollistamiseen.

Suomessa tulisi tehdä myös big data -alan teknologiakartoituksia ja olisi tunnistettava oleellisimpia teknologioita, malleja sekä standardeja. Samalla tulisi kehittää erilaisten big dataan liittyvien määritelmien ja termien yhtenäisyyttä.

Yksityisyydensuojan ja käyttäjästä kerättävän tiedon kannalta käytössä on erilaisia teknisiä työkaluja sekä hallinta- ja käyttömenetelmiä. Menetelmät ja välineet sopivat erilaisiin tarkoituksiin ja niillä on omat etunsa ja haittansa.<sup>32</sup> Menetelmiin ja välineisiin kuuluvat muun muassa:

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

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

Syntyvä data halutaan järjestelmätasolla tallettaa big dataa käsittelevään järjestelmään. Tyypillisesti tämä edellyttää riittävää tiedonsiirtokapasiteettia, jonka puuttuessa dataa ei ehkä ennen ole pystytty tallentamaan tai siitä on talletettu vain osa. Tiedon varastointia varten tarvitaan tallennuskapasiteettia. Big dataa voidaan ja joskus pitääkin yrittää analysoida nopeasti ja lähinnä käyttäen lähihistoriaa, mutta usein sitä halutaan säilyttää pitkiäkin aikoja ja suuria määriä. Tällöin tarvitaan fyysistä tallennuskapasiteettia.

Tiedon analysointiin tarvitaan laskentakapasiteettia. Yksittäisen tietokoneen datan käsittelykapasiteetin nousu ei ole enää riittävä tähän, joten erityisen suurten datamassojen

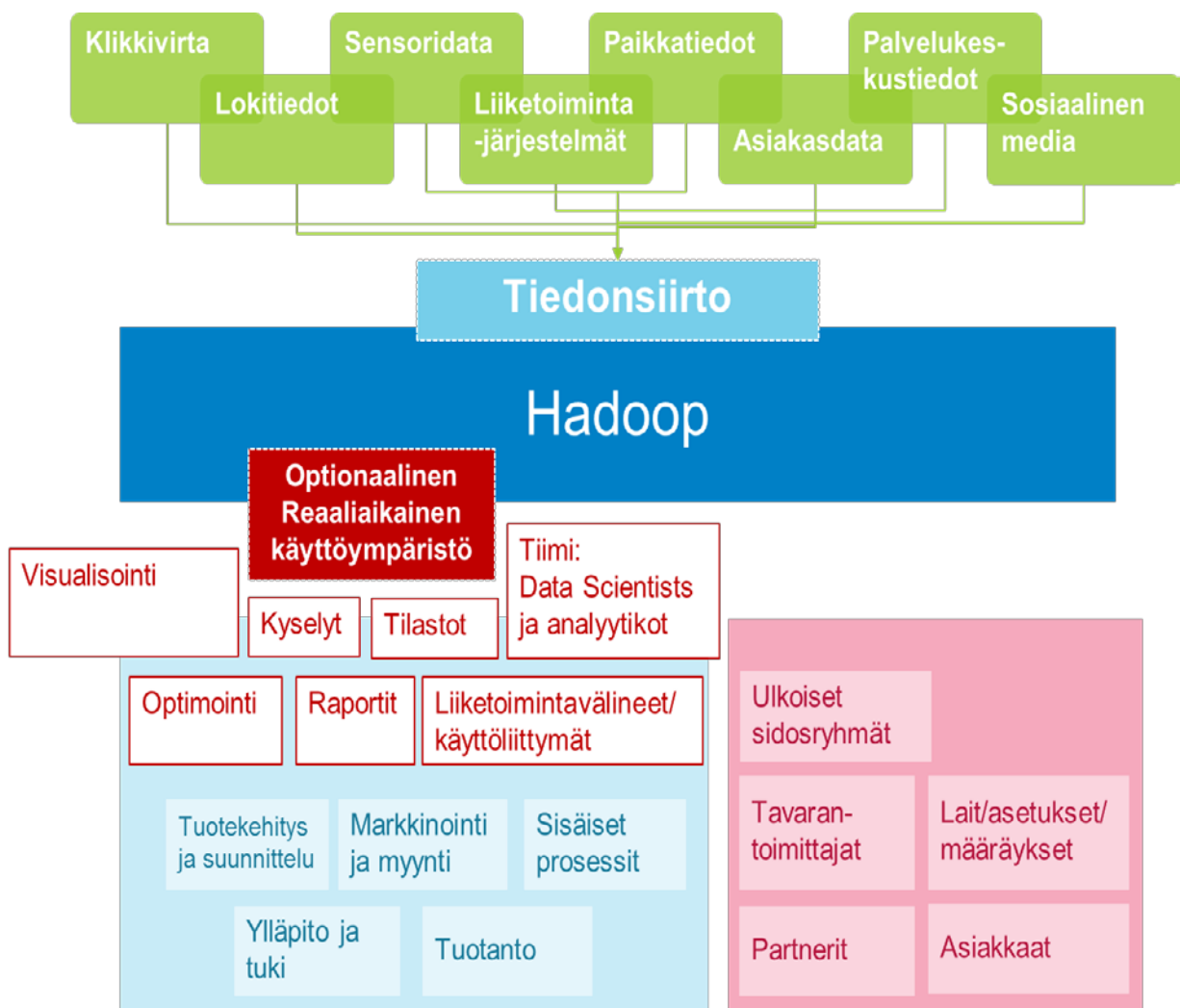
<sup>31</sup> NIST Big Data Working Group. <http://bigdatawg.nist.gov/home.php>

<sup>32</sup> Executive Office of the President (2014). Big Data and Privacy: a technological perspective. [http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast\\_big\\_data\\_and\\_privacy\\_-\\_may\\_2014.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf)

suhteen on siirrytty käyttämään hajautettua ja samanaikaista laskentaa. Laskentatehtävät jaetaan vaiheisiin, jotka valikoivat ja jalostavat dataa seuraavaa vaihetta varten. Näitä vaiheita suoritetaan hajautetusti eri tietokoneissa. Myöhemmät vaiheet voivat yhdistellä ja summata dataa tässä hajautetussa laskenta-arkkitehtuurissa siten, että laskennan edetessä ja välitulosten pienentyessä ja yhdistyessä laskennan lopputulos saadaan muodostettua.

Mikäli datan suuruus liittyy pikemminkin sen monimuotoisuuteen kuin sen määrään, niin tarvittavat ratkaisut liittyvät enemmän dataintegraatioon ja ovat enemmän ohjelmistokuin laitteistoteknisiä.

**Kuva 6. Esimerkki organisaation big data -arkkitehtuurista**



Big data -tekniikan ytimessä ovat yleisimmin avoimen lähdekoodin ratkaisut, kuten esimerkiksi Apache Software Foundationin Hadoop-projekti, sekä siihen suoraan tai välillisesti liittyvät projektit ja työkalut, kuten Hive, Storm, Spark ja Kafka. Hive mahdollistaa kyselyjen tekemisen suuriin Hadoopiin tallennettuihin tietoaisteihin ja Storm puolestaan reaaliaikaiseen Hadoop-laskennan toteuttamisen. Spark-tekniikka tarjoaa Hadoop-



pin alkuperäistä nk. map-reduce -mallia monipuolisemman laskennallisen ympäristön ja on käytettävissä myös Hadoopiin liitettynä.

Moderni big data -laskenta perustuu hajautusta ja samanaikaisuutta hyödyntävään monivaiheiseen laskentaan. Hadoop on käytännössä merkittävin ja laajasti käytetty tätä periaatetta toteuttava avoimen lähdekoodin järjestelmä. Hadoop-järjestelmä on sinänsä hankalakäyttöinen ilman apuvälineitä, joten sitä täydentämään on kehitetty ja kehitetään erilaisia työkaluja.

Nykyään suuri osa big datan käsittelystä tapahtuu ohjelmoimalla operaatiot suoraan Hadoopia käyttäen, mutta teknologiakehityksen painopiste on siirtynyt mm. tässä mainittuihin projekteihin, ja korkeamman tason työkalujen painoarvo tulee jatkossa kasvamaan merkittävästi. Toisaalta big data -laskennan tehokkuus liittyy siihen, kuinka data on järjestetty, sekä datan organisaatiota vastaavien algoritmien suunnitteluun, joka säilyttää tarpeen suoralle ohjelmoinnille.

Big datan käsittelyyn reaaliaikaisissa ympäristöissä liittyy usein NoSQL-tietokantojen, kuten MongoDB:n hyötykäyttäminen. Kokonaisvaltaiset big data -ympäristöt rakentuvat yleisesti hybridiympäristöistä, joissa on mukana myös perinteisiä, esim. SQL-pohjaisia ympäristöjä, sekä uutta teknologiaa. Näiden ympärille on olemassa useita kaupallisia uusia ja perinteisiä ratkaisuja, joilla tietoa saadaan raportoitua, sekä esitettyä visuaalisesti. Raportointi- ja visualisointityökalujen avulla big datan käsittelyyn ja hyödyntämiseen ei tarvita ohjelmointiosaamista.

## 2.6 Datan avoimuus ja yhdistely

Yksi tapa lähestyä big datan käytettävyyttä ja avoimuutta on esitetty kuvassa 7.<sup>33</sup> Potentiaalisten hyötyjen kannalta mielenkiintoisimpia alueita ovat big datan ja avoimen datan risteyskohdat. Tämä voi näkyä myös suoraan liiketoiminnassa. Esimerkiksi Aalto yliopiston tutkimuksessa vuonna 2012 havaittiin, että vapautuva julkinen data on tärkeää pk-yritysten liikevaihdon kasvun kannalta<sup>34</sup>. Julkisten tietoaaineistojen avaaminen vapaaseen uudelleenkäyttöön on olennaista myös big datan hyödyntämisen kannalta.

Data rinnastetaan useissa puheissa luonnonvaroihin tai hyödykkeisiin, jolla on arvoa myös kauppatavarana<sup>35</sup>. Datan kerääminen ja koostaminen on synnyttänyt ja synnyttää edelleen uusia toimijoita markkinoille ja uusia markkinoita. Dataa voidaan myydä joko yksittäisten palveluiden kautta tai koostettuna yleisille datan markkinapaikoille. Kysyntä ja tarjonta määrittävät ennen pitkää erilaisen datan arvon.

Jotta big datasta saadaan enemmän hyötyä irti, täytyy raaka-aineen eli datan itsensä, olla mahdollisimman laajasti avointa ja helposti käytettävissä. Erilaisten tietoaaineistojen avaaminen ja niiden yhdistely voi mahdollistaa uusia palveluinnovaatioita. Tietoaaineisto-

<sup>33</sup> Mukaillen: Gurin (2014), The relationship between big data and open data. <http://www.theguardian.com/public-leaders-network/2014/apr/15/big-data-open-data-transform-government>

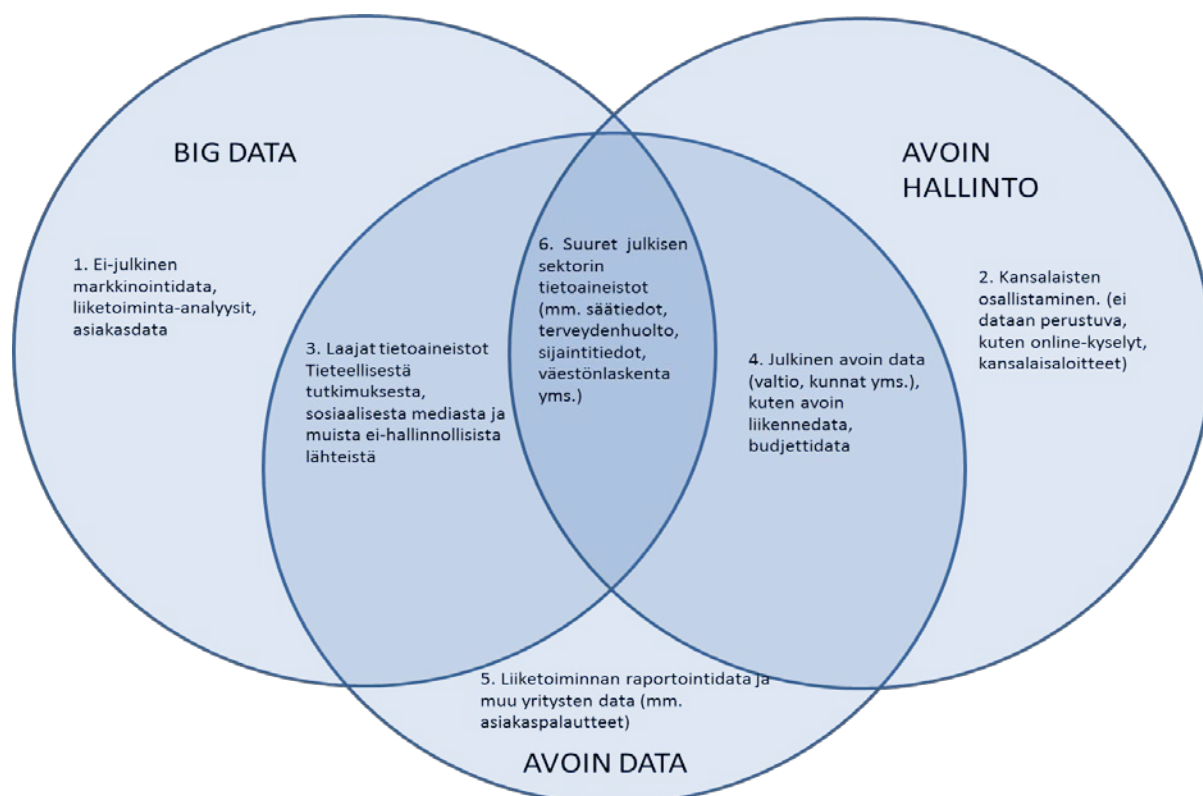
<sup>34</sup> Kaipiainen-Heiskanen (2012), Julkisen tiedon vapautuminen kasvattaa pk-yritysten liikevaihtoa. <http://pienyrittyskeskus.aalto.fi/fi/info/ajankohtaista/view/2012-06-19/>

<sup>35</sup> Laney (2013), Infonomics: The new economics of information. <http://www.ft.com/cms/s/0/205ddf5c-1bf0-11e3-b678-00144feab7de.html#axzz35XpcjdQV>

jen avaaja ei useimmiten voi tietää, miten avattua tietoaineistoa käytetään ja mihin sitä mahdollisesti yhdistellään.

Julkisen tiedon tulee olla mahdollisimman laajasti avointa ja maksutonta. Julkisen tiedon hyödyntäminen on kirjattu tavoitteeksi myös Jyrki Kataisen hallituksen hallitusohjelmaan. Tietovarantoja avataan osana valtiovarainministeriön vetämää avoimen tiedon ohjelmaa<sup>36</sup>.

### Kuva 7. Big datan, avoimen datan ja avoimen hallinnon yhteydet



Sekä EU- että kansallisen tason lainsäädäntö ja politiikka ohjaavat 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 vauhdittaakseen oman erityisosaamisalueensa kehittymistä sekä hankkiakseen hyvää mainetta. Organisaatioiden onkin syytä ottaa mukaan kokonaisvaltaiseen data-strategiaansa myös avoin data ja sen erilaiset mahdollisuudet.

Nykyaikainen suhtautuminen datan avaamiseen ja tiedon jakamisen kulttuurin kehitys luovat myös kansallista pääomaa ja auttavat Suomea toimimaan kansainvälisenä esimerkkimaana. Kokemukset dataan avaamiseen liittyvistä päätöksistä, käytännön toteutuksista ja vaikutuksista liiketoimintakenttään voivat lähitulevaisuudessa olla haluttua pääomaa. Parhaiden käytäntöjen tuotteistaminen voi luoda myös hyvää tukea suomalaisen osaamisen vientiin.

<sup>36</sup> Valtiovarainministeriön avoimen tiedon ohjelma. [http://www.vm.fi/vm/fi/05\\_hankkeet/02381\\_avoin\\_tieto/index.jsp](http://www.vm.fi/vm/fi/05_hankkeet/02381_avoin_tieto/index.jsp)

Avoimen tiedon tulee olla muodossa, jossa sen hyödyntäminen on mahdollisimman helppoa. Esimerkiksi erilaisten julkisten tietoaisteistojen koneluettavia standardimuotoja tulee edistää. Tärkeä on huomata myös suuria tietoaisteistoja kuvailevan metatiedon muodot. Koneluettavat, standardoidut metatietomallit mahdollistavat relevanttien aisteistojen löytämisen ja tehokkaan jatkokäytön.

Datan käytettävyyteen vaikuttavat myös sen hallintamenetelmät ja omistajuus. Eriluonteisen tiedonhallinta tarvitsee erilaisia käytäntöjä. Henkilötietojen ja tiedon keräämisen ja käytön periaatteiden tulisi olla läpinäkyvästi ja ymmärrettävästi esitetty ja hyvin kansalaisten tiedossa.

## 2.7 Sääntely ja tietosuojaja

Tietosuojaja on noussut keskeiseksi kysymykseksi keskusteltaessa tietoaisteistojen keräämistä, säilyttämistä ja hyödyntämistä koskevasta lainsäädännöstä. 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 tietosuojaja-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. Tietosuojaja-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 suojaa ja julkisuutta koskevan sääntelyn lisäksi big datan käsittelyyn voi vaikuttaa **immateriaalioikeudellinen sääntely**. Erilaiset tietokannat ja niihin liittyvät ohjelmistot saattavat esimerkiksi nauttia tekijänoikeudellista suojaa, joka voi rajoittaa esimerkiksi tekstimuotoisen tiedon tutkimuskäyttöön suuntautuvaa louhintaa. Lisäksi tietoihin ja niiden käyttöön saattaa liittyä sopimusoikeudellisia ja muita kaupallisia velvoitteita, jotka tietojen käsittelijän tulee huomioida. Uutta tutkimusta mahdollistavia poikkeamia tekijänoikeuksiin täytyy pohtia vakavasti. Tietoaineistojen käytettävyyttä voivat lisätä myös erilaiset lisenssi- ja IPR -sopimusmallit.

Digitaalisten aineistojen hyödyntäminen vaatii tuekseen tarkoituksenmukaista sääntelyä. Tarvitaan yhteiskunnallista ja juridista ymmärrystä siitä, miten aineistoihin liittyvää lainsäädäntöä voidaan kehittää tavalla, joka vie eteenpäin toivottuja tapahtumakulkuja. Muun muassa digitaalisesta jäljittämisestä pitää käydä avointa keskustelua. Sen hyödyistä ja haitoista on oltava tutkimuksellisesti perusteltua näyttöä.

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.

## 2.8 Yhteistyö ja datan vaihto

Tietoyhteiskunnasta verkostoyhteiskuntaan siirtyminen on lisännyt yritysten tiedon hyödyntämisen tarvetta merkittävästi. Jos aiemmin yritysten kilpailuetuna pidettiin suljettuja organisaatioita, jotka pyrkivät suojelemaan kaikkea hankkimaansa tietoa, uudempien liiketoimintamallien etuina on nähty siirtyminen avoimempien, yhteistyötä tekevien organisaatioiden suuntaan<sup>37</sup>. Kilpailukykyiset yritykset ovat riippuvaisia myös yhteistyöverkostoista, alihankkijoistaan ja tiedon jakamisesta. Toisaalta on tehty myös huomioita siitä, miten merkittävä osuus uusista innovaatioista tulee olemassa olevan nykyisen kilpailun ulkopuolelta<sup>38</sup>.

<sup>37</sup> Wikinomics-ajattelusta ks. Tapscott & Williams (2006), Wikinomics: How Mass Collaboration Changes Everything.; Liiketoimintaympäristön muutoksista ks. Nelson (2010), Business Intelligence 2.0: Are we there yet? <http://support.sas.com/resources/papers/proceedings10/040-2010.pdf>

<sup>38</sup> Sydänmaanlakka (2009), Jatkuva uudistuminen – luovuuden ja innovatiivisuuden johtaminen; Nonaka, Toyama & Hirata (2008), Managing Flow – a Process Theory of the Knowledge-Based Firm.

Nopeasti muuttuvassa maailmassa liiketoimintaympäristön muutoksiin täytyy reagoida. Suomalaisen yrityskehityksen täytyy seurata, omaksua ja tarvittaessa myös nopeasti kopioida dataa ja analytiikkaa onnistuneesti hyödyntäneitä uudenlaisia liiketoimintamalleja muualta maailmalta. Oman liiketoimintatietonsa hyödyntämistä suunnittelevat yritykset voisivat tehokkaammin valjastaa myös pienempien yritysten innovaatiokyvyn ja ketteryyden tukemaan omaa liiketoimintaansa. Pienet yritykset voivat saada yhteistyön kautta uskottavuudelle tärkeitä referenssejä.

*”Suurimmat ongelmat liittyvät siihen, miten johto ymmärtää omistetun tiedon ja kääntää sen liiketoiminnan parhaaksi”*

Yritysten välisessä verkostoitumisessa on myös haasteita. Suomessa pienempien innovatiivisten yritysten ja isoja datamäärät omaavien suuryritysten väliset systemaattiset yhteistyömallit ovat suhteellisen harvinaisia. Mahdollistaakseen innovaatiopotentialin tehokkaamman hyödyntämisen, voisivat suuria määriä dataa liiketoiminnassaan keräävät yritykset avata dataansa avoimen kehitysalustan tai muiden rajapintojen kautta yrityksille, jotka kehittävät datasta uusia palveluita. Suomen laaja start-up -kenttä, esimerkiksi AaltoES, Protomo ja LutES tulee kytkeä tiiviisti mukaan big data -ratkaisujen ja palvelujen kehitykseen.

On todennäköistä, että tulevaisuudessa yritysten on teknisesti entistäkin helpompaa avata myös kokonaisia big data -varantoja ulkoisille analyyseille, hyödyntäen tähän tarkoitukseen kehitettäviä standardeja analyysin kehitysympäristöjä.

Kansainvälisessä kehityksessä Saksa on avainasemassa teollisen internetin tulevissa ratkaisuihin. Suomalaisilla yrityksillä ja niiden osaamisella on paljon erilaisia mahdollisuuksia näiden globaalisti johtavien yritysten big data -kehityksessä. Useissa maissa on yhteistyötä edistämään perustettu big data -laboratorioita, -tutkimuskeskuksia ja -verkostoja, kuten esimerkiksi Saksan Smart Data Innovation Lab ja Singaporen Social Analytics -yhteistyökonsortio<sup>39</sup>. General Electric puolestaan tarjoaa kumppaneilleen tiloja kehittääkseen datan hyödyntämistä liiketoiminnassaan. Suomalaisyritysten on kehitettävä tapoja datan jakamiseen ja lisäarvon tuottamiseen yhteistyössä, jotta yritysten sovellusosaaminen ja yleinen alan kehitys vauhdittuisi.

<sup>39</sup> iDA: Social Analytics (SA) for Business Enterprises Call-for-Collaboration (CFC). <http://www.ida.gov.sg/Collaboration-and-Initiatives/Collaboration-Opportunities/Store/Social-Analytics-SA-for-Business-Enterprises-Call-for-Collaboration-CFC>

Saksassa yhteistyötä tehdään muun muassa suuret ja pienet yritykset sekä tutkimuspuolen yhteen tuovan Smart Data Innovation Labin kautta.

Lähde: [http://www.kit.edu/kit/english/pi\\_2014\\_14408.php](http://www.kit.edu/kit/english/pi_2014_14408.php)  
<http://www.sdil.de/de/>

Suomessa Tiedon vetämä, strategisen huippuosaamiskeskittymä DIGILEn Data to Intelligence -ohjelma (D2I) kokoaa yhteen suuria yrityksiä, pk-yrityksiä ja tutkimuspuolta löytämään yhdessä ratkaisuja big datan hyödyntämiseen.

Lähde: <http://www.datatointelligence.fi/>

Datan jakaminen voi olla teknisesti haasteellista. Dataa saattaa olla niin paljon ja sitä voi syntyä niin suurella nopeudella, että sitä ei voi siirtää fyysisesti toisen organisaation käsiteltäväksi. Käytännössä tiedon analysointia varten täytyy siirtää prosessoivaa koodia toisen organisaation sisälle. Tämä on ongelma, joka vaatii sekä teknistä että tietoturvallista ratkaisua. Tulevaisuudessa oma roolinsa voi olla myös dataoperaattoreilla, jotka huolehtivat datan analyysistä ja resursoinnista yhteistyössä osapuolten kanssa. Big datan käsitelyyn voidaan valjastaa myös kiinteiden organisaatioiden ulkopuolista prosessointivoimaa muun muassa joukkoistamalla ja avaamalla tietovarantoja innovointiin avoimiin tietoverkkoihin.

Samalla tulee huomioida viranomais- ja yritystiedon yhdistämisen mahdollisuudet. Yrityksillä ja eri palveluntarjoajilla hallussa oleva tieto saattaa olla julkisen hallinnon ja viranomaisten toiminnan kannalta kullannarvoista. Vaikkapa liikennevirtojen suunnittelussa datalähteiden yhdistäminen tarjoaa huomattavia mahdollisuuksia. Sama pätee luonnollisesti myös toisinpäin; julkiselta sektorilta löytyvä tieto mahdollistaa moni uusia palveluita.

## 2.9 Kokeilut ja T&K -rahoituksen suuntaaminen

Nopeaan kaupallistumiseen ja kansainväliseen, skaalautuvaan liiketoimintaan tähtääville big data -kokeiluille tulisi saada riittäviä resursseja ja rahoitusta, muun muassa kaupallistamiseen ja tuotekehitykseen. T&K -panostuksia tulisi kohdentaa myös big dataan liittyväälle tutkimukselle. Esimerkiksi EU:n Horisontti 2020 -ohjelma tarjoaa rahoitukseen mahdollisuuksia.

Organisaatioiden ulkopuoliset tietovarannot ovat aina organisaation omasta koosta riippumatta sen sisäisiä tietovarantoja suurempia, joten organisaatorajat ylittävään datan yhdistelyyn perustuvat nopeat (myös epäonnistumisen suhteen nopeat) ja skaalautuvat monen toimijan ketterät kokeilut todennäköisesti tuovat innovatiivisemmat lopputulokset. Yritysten yhteiset ja tutkimuksen kanssa toteutettavat kokeilut voivat edistää uutta innovoinnin kulttuuria.

Erilaisia kokeiluja voidaan tehdä myös muun muassa erilaisten laajojen tutkimushankkeiden yhteydessä. Kokeiluja on käynnistymässä jo muun muassa käyttäjälähtöiseen datan hyödyntämiseen liittyen.

Ylipäänsä big dataan liittyviä ja kohdentuvia T&K -investointeja täytyy tehdä, mikäli Suomi pyrkii pysymään kehityksessä mukana. Big data -kehitykselle on useissa maissa osoitettu huomattavia investointeja valtion taholta. Suomessa julkisten toimijoiden tulee osittain myös paikata yksityisten pääomasijoitusten puutetta.

### Horisontti 2020

EU:n Horisontti 2020 -ohjelma tarjoaa big dataan liittyviä rahoitusmahdollisuuksia. Esimerkiksi 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.

Lähde: <http://ec.europa.eu/programmes/horizon2020/en/h2020-section/information-and-communication-technologies> , [http://ec.europa.eu/research/participants/data/ref/h2020/wp/2014\\_2015/main/h2020-wp1415-leit-ict\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/main/h2020-wp1415-leit-ict_en.pdf)

### Kokeiluja tutkimushankkeessa: Digital Health Revolution

Tekes rahoittaa noin neljällä miljoonalla eurolla kuuden yliopiston ja tutkimuslaitoksen monitieteistä Digital Health Revolution -tutkimushanketta. Kokeilut ovat hankkeessa vahvassa osassa:

"Hankkeen lähtökohtana on, että yksilö voi hyödyntää omia tietojaan itse valitsemis- ja käyttökohteissa sekä kontrolloida tietojensa käyttöä. Hankkeen tavoitteena on yhdistää systemaattisesti eri lähteistä saatavaa tietoa, mukaan lukien yksilön genomitieto, terveys- ja hyvinvointiseurantatieto sekä arkikäyttötietoon liittyvä tieto eli ihmisen digitaalinen jalanjälki. Tavoitteena on älykkäästi jalostaa yksilöllinen terveys-tieto yhteiskunnan ja yksilön hyödyksi uusina palveluratkaisuina.

Hankkeessa tullaan rakentamaan henkilökohtaiseen tietoon perustuvia käyttäjälähtöisiä, ennaltaehkäiseviä palveluratkaisuja palvelupilottien avulla. Hankkeella pyritään perinpohjaiseen asenneilmapiirin muutokseen liittyen digitaalisen tiedon hyödyntämisen tapoihin. Siksi tekemisen ytimessä on MyData.fi -yhteisön rakentaminen, mikä samalla myös mahdollistaa uuden liiketoiminnan ja uusien palveluratkaisujen kehittämisen eettisiä ja lainmukaisia säännöksiä noudattaen.

Pitkällä aikajänteellä hankkeen tulokset mahdollistavat henkilökohtaiseen tietoon perustuvien ennaltaehkäisevien palveluratkaisujen integroitumisen myös osaksi julkista terveydenhoitoa ja tulevaisuuden palvelurakennetta."

Lähde: <http://www.tekes.fi/nyt/uutiset-2014/tekes-rahoittaa-kolme-uutta-strategista-tutkimusavausta/>

## 3. Sovellusalueet ja niiden potentiaali Suomessa

Työryhmä on tunnistanut ja priorisoinut Suomen kannalta olennaisia laajojen tietoa-aineistojen hyödyntämisen sovellusalueita. Tunnistetuissa sovellusalueissa Suomella on erityistä osaamista tai niistä odotettavat hyödyt ovat talouden ja koko yhteiskunnan kannalta niin merkittäviä, että big datan hyödyntämiskäytännöt tulisi saada mahdollisimman pian käyttöön. Monella muulla toimialalla datan käyttö voi olla jo hyvinkin edistynyt, esimerkiksi kauppaa ja finanssialat.

Raportissa priorisoidut sovellusalueet ovat kuitenkin toimialoja, jotka samalla tarjoavat Suomelle potentiaalia tulevaisuudessa. Niillä tarvitaan kuitenkin erityisiä kehitystoimia odotettujen hyötyjen saavuttamiseksi. Sovellusalueiden lisäksi on tunnistettu läpileikkäviä teemoja, joihin myös liittyy kasvuodotuksia.

### 3.1 Terveys

Suomesta löytyy maailmanlaajuisesti katsoen poikkeuksellisen laadukkaita ja kattavia terveystietokantoja, muun muassa geenitietoihin liittyen. Tämän vahvuuden vastuullinen hyödyntäminen big data -ratkaisuisa voisi tarjota Suomelle merkittävän, globaalin kilpailukykytekijän. Suomessa olisi innovaatiotilasta erityisesti ikääntyvälle väestönosalle suunnatuille ennaltaehkäisevän terveydenhuollon ja itsehoidon palveluille. OECD:n Suomea koskevan raportin mukaan<sup>40</sup> väestön ikääntymisen aiheuttamat rakenteelliset terveydenhuollon kustannuspaineet tulevat lähivuosina ja myös tulevaisuudessa olemaan Suomessa merkittävämpi ongelma kuin useimmissa muissa OECD-maissa.

Terveydenhuollon kenttä liittyy big dataan lukuisilla eri tavoilla ja Suomessa on alan osaamista esimerkiksi bioinformatiikan ja molekyyli lääketieteen alalla. Toinen tärkeä alue liittyy kansalaisten omatoimiseen terveydenhoitoon ja hyvinvointiin liittyvien itsemitausteknologioiden kehitykseen, jossa suomalaisyrityksissä on osaamista kymmenien vuosien ajalta. Terveysdatan laajemman saatavuuden ennustetaan muuttavan ihmisten käsitystä itsestään ja omasta terveydestään. Elämäntavoista ja geenitiedosta kertyvän yhdistetyn big datan avulla voidaan mahdollisesti tulevaisuudessa ehkäistä tai viivyttää sairastumista ja edistää terveyttä.

Koko sosiaali- ja terveystoimialan kannalta suurten tietoaaineistojen analyysi tarjoaa mahdollisuuden tarkemman tilannekuvan muodostamiseen ja palveluiden kohdentamiseen. Tämän mahdollistavat muun muassa alan erinomaiset tietovarannot ja niiden jatkokehittäminen. Lisäksi big data mahdollistaa terveydenhuollolle uudenlaisen oppimisen, kun kerrytetty ja analysoitu data tukee lääkärin ja muun hoitohenkilökunnan työtä.

Big datalta odotetaan terveydenhuollon laadun parantumista ja säästöjä. Data-analyysin odotetaan kehittyvän tavalla, joka auttaa yksittäisen kansalaisen terveystietojen ennu-

*”Suuret tietoaaineistot parantavat omahoitoa, ennaltaehkäisevää toimintaa, valvontaa sekä tukevat ammattilaisten työtä”*

tamisessa erilaisten datavirtojen avulla sekä sairauksien ehkäisyssä. Terveystietojen hyötykäyttöön liittyy kuitenkin yksityisyyteen, sosiaalisiin ja eettisiin näkökulmiin liittyviä kysy-

myksiä, jotka on otettava vakavasti. Ilman näiden tekijöiden hyödyntämistä suotuisat kehityskulut tuskin ovat mahdollisia.

Yksi tapa on määritellä vastuullisen aineiston käytön puitteita yhteistyössä eri alojen osaajien kanssa. Suomessa tulisikin 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.

<sup>40</sup> OECD (2014) Economic Surveys: Finland. [http://www.oecd.org/eco/surveys/Overview\\_Finland\\_2014.pdf](http://www.oecd.org/eco/surveys/Overview_Finland_2014.pdf)



Nykyisessä taloudellisessa tilanteessa, varsinkin kuntien osalta, on säästöjen saavuttaminen uuden teknologian avulla mahdollisimman hyvä palvelutaso säilyttäen avainasemassa. Erilaisia dataan perustuvien omahoitomenetelmien käyttöönottoa tulee samalla tukea. On myös tärkeää, että suomalaiset tuntevat kansainvälisen terveystieteen big data -kentän ja ovat mukana vaikuttamassa sen kehitykseen.

### **Suomalaisessa syöpätutkimuksessa sovelletaan koneellista mallintamista**

HIITissä tehdään urauurtavaa tutkimusta valtavien tietoaaineistojen hyödyntämisessä muun muassa tutkittaessa syöpälääkkeiden vaikutuksista soluihin. Genominlaajuisten monimutkaisten vaikutusmekanismien tutkimiseksi on pitänyt yhdistää useita tietokantoja toisiinsa sekä käyttää innovatiivisia tilastomatemaattisia ratkaisuja. Koneellisella mallintamisella voidaan löytää parhaiten sopivia lääkkeitä kunkin genomirakenteen tai sairastetun sairaustyyppin mukaisesti.

Lähde: Comprehensive data-driven analysis of the impact of chemoinformatic structure on the genome-wide biological response profiles of cancer cells to 1159 drugs, Kaski ym.

[http://research.ics.aalto.fi/publications/bibdb2012/public\\_pdfs/Khan12bmcbioinformatics.pdf](http://research.ics.aalto.fi/publications/bibdb2012/public_pdfs/Khan12bmcbioinformatics.pdf)

### **Laskennalliset mallit ennustavat lääkaineiden vaikutuksia**

Laskennallinen täsmälääketiede kehittää laskennallisia malleja, joilla voidaan ennustaa lääkaineiden tehoa ja vaikutusta soluissa niiden geeniprofiilien pohjalta. Suomalainen monitieteinen tutkijaryhmä voitti tietoanalyttiseen lääkeherkkyyden ennustamisen menetelmiä koskevan kilpailun, jonka Yhdysvaltain kansallinen syöpäinstituutti ja DREAM-yhteisö (Dialogue on Reverse Engineering Assessment and Methods) olivat järjestäneet. Parhaan tuloksen antoivat epälineaariset mallit ja eri mittaustekniikoiden painottaminen, käyttämällä koneellisen oppimisen metodeja. Tutkimus antaa vahvan perustan tuleville tutkimuksille mallien kehittämisestä potilastiedoista

Lähteet: "A community effort to assess and improve drug sensitivity prediction algorithms", <http://www.nature.com/nbt/journal/vaop/ncurrent/full/nbt.2877.html>, lisäksi Suomen molekyyllilääketieteen instituutti FIMM: <https://www.fimm.fi/fi/tapahtuma/1401965486>

## **3.2 Älykkäät infrastruktuurit**

Datan merkityksen lisääntyminen heijastuu myös yhdyskuntasuunnitteluun ja energianjakeluun. Esimerkiksi kaupungeissa älyverkkojen mahdollistamat uudet energianjakelun ja asumisen ratkaisut tulevat olemaan arkipäivää. Älykkäät verkot tarvitsevat toimivan verkkoinfran lisäksi toimiakseen myös älysovelluksia, joissa hyödynnetään avoimia datamassoja ja tietovarantojen infrastruktuureja.

Suomessa älykaupunki-ajattelun kehityskärjessä kulkevat useat tutut yritykset, kuten Siemens, ABB, Schneider, sekä Helen, Elenia ja Fortum. Asumisen innovatiivisiin palveluratkaisuihin kuuluvat muun muassa kotiin asennettavat anturit, jotka mittaavat sisäilman laatua. Älykaupunkien ja älykkäiden verkkojen kasvualustana ovat data, valtavasti kas-

vanut laskentateho, anturiteknologian kehittyminen, sekä automatisoinnin mukanaan tuomat mahdollisuudet.

Energiatehokkuuden lisäksi myös resurssien käyttö on älykaupunki-ajattelussa entistäkin tehokkaampaa, ja synnyttää ympärilleen uusia älypalveluita. Verkkojen avulla kaupunki-en älytalot tuottavat valtavat määrät dataa, jota voidaan käyttää uusien älykaupunki-palvelujen rakennusaineena.<sup>41</sup>

Älykkäät sähköverkot mahdollistavat myös sähkön myymisen takaisin verkkoon, jolloin erilaisia sähkölaitteita on kytketty älykkäästi toisiinsa ja varastoimaan ylimääräistä energiaa. Älyverkkojen merkitys korostuu kylmissä olosuhteissa, joissa anturit 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 aikaan, jolloin uusiutuvaa energiaa on saatavilla.

Älykkäiden sähköverkkojen data kerätään ja hallinnoidaan. MDMS-hallintajärjestelmällä (Meter Data Management System). Antureista ja mittareista saatava tieto voidaan ohjata myös uusien palvelujen ja sovellusten kehittämiseen. On ennustettu, että mitattavan datan hallintajärjestelmien markkinat kasvavat yli 300 % vuonna 2014<sup>42</sup>.

Suunniteltaessa ja toteutettaessa älykkäitä verkkoja ja älykaupunki-teknologioita hyödyntäviä uusia palvelukonsepteja, on tärkeää huomioida alusta asti ketterän ja kommunikoivan yhteistyön merkitys. Yhteistyötä tehdään usein projektia sponsoroivan kunnan tai kaupungin, palveluita käyttävien kansalaisryhmien ja teknisten palveluntarjoajien välillä. Kuten monen muun uuden teknologian soveltamisessa, tulee älykkäämmän kaupungin infraa hyödyntävien palveluiden toteutuksessa lähteä liikkeelle ratkaisujen käytännön operoinnista sekä asiakkaiden ja yksittäisten kaupunkien erilaisisten tarpeiden sovittamisesta uuteen teknologiaan.

Kaupunki-infran uudenlaisia älykkäitä palveluita pilotoivat ja skaalaavat projektit tarvitsevat siis jatkuvaa tukea niin kansalaisilta, poliittisilta päättäjiltä kuin kaikilta projektin kustannuksista vastaavalta tahoilta.

---

<sup>41</sup> Liikenne- ja viestintäministeriö. Älykäs kaupunki – Smart City: Katsaus fiksuihin palveluihin ja mahdollisuuksiin.

[http://www.lvm.fi/c/document\\_library/get\\_file?folderId=3082174&name=DLFE-23659.pdf&title=Julkaisu%2012-2014](http://www.lvm.fi/c/document_library/get_file?folderId=3082174&name=DLFE-23659.pdf&title=Julkaisu%2012-2014)

<sup>42</sup> greentechgrid (2010), Meter Data Management Market to Grow 300% by 2014, According to GTM Research.

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

### **Mobiilidatatilastot kaupunkisuunnittelun apuna**

Aiemmin staattisten asutustilastojen pohjalta tehtyä kaupunki- ja tilasuunnittelua voidaan tehostaa merkittävästi käyttämällä tukena matkaviestinten verkkodatatilastoja. Dynaamiset mallit ovat tärkeässä asemassa, sillä sekä ikä- että palvelurakenteiden muutosten sekä väestön liikkeiden arviointiin tarvitaan yhä täsmällisempiä ennusteita ja malleja. Näistä malleista on apua muun muassa tie- ja liikennesuunnittelussa, maankäytön suunnittelussa, maisemoinnissa sekä puisto- ja viheraluesuunnittelussa.

Japanin suurin mobiilipalvelujen tarjoaja NTT Docomo teki yhteistyössä Tokion yliopiston kanssa tutkimuksen 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. Tästä tutkimuksesta saatua tietoa voitiin käyttää perustietona 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 tietoihin.

Lähde: 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](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 uusia mahdollisuuksia. Esimerkiksi uusien, dataan perustuvien tutkimusmenetelmien kehitys tuo tutkijoille huomattavasti parantuneita mahdollisuuksia tutkia erilaisia asioita ja löytää ratkaisuja ongelmiin. Muun muassa erilaisen mobiilidatan, trendidatan sekä sosiaalisesta mediasta saatavan datan määrän kasvu tuo tutkijoiden ulottuville aivan uudenlaisia aineistoja. Erimerkiksi ihmisten todellista käyttäytymistä kuvaavat verkkokäyttö- ja paikkatietoaineistot mahdollistavat merkittävien lisätietojen saamisen ja siten täydentävät (tai jopa korvaavat) perinteisiä kyselytutkimuksia, joilla tutkitaan muun muassa ihmisten käyttäytymisaikomuksia. Uusia mahdollisuuksia luo muun muassa Googlen keräämän 10-vuotisen kyselytrendidatan käyttö erilaisissa ennustemalleissa. <sup>43</sup>

Luonnontieteellisillä aloilla kuten biologiassa, lääketieteessä, tähtitieteessä ja fysiikassa on jo pitkät perinteet datalähtöisestä ja supertietokoneita hyödyntävästä tutkimuksesta.

<sup>43</sup> Varian (2013), Big data: New Tricks for Econometrics. <http://people.ischool.berkeley.edu/~hal/Papers/2013/ml.pdf>

Myös tietyillä taloustieteen alueilla kuten rahoituksessa ja ekonometriassa data on en- tuudestaan ollut jo hyvin keskeisessä roolissa tutkimuksessa.

Tekstinlouhintamenetelmien jatkuva kehitys on mahdollistanut uusia tutkimusalueita eri- tyisesti humanististen ja yhteiskuntatieteiden alueilla. Digitaalisessa markkinoinnissa tut- kitaan muun muassa kuluttajien sentimenttejä (positiivisia tai negatiivisia tuntemuksia) tuotteita tai yrityksiä kohtaan ja vastaavaa analyysiä käytetään jopa kokonaisten kansa- kuntien tuntemusten kartoittamiseen reaaliajassa.<sup>44</sup>

### **Taloustuutisten sentimenttianalyysiä tutkimuksessa**

Suomessa sentimenttianalyysiä käytetään ja kehitetään muun muassa Aalto-yliopiston kauppakorkeakoulun tutkimushankkeessa, jossa selvitetään miten online-taloustuutisia, sijoitusaiheisia blogeja –ja foorumeita voidaan analysoida ohjelmallisesti tavalla, joka mahdollistaa vallitsevien markkinasentimenttien ja trendien tunnistamisen tarkastelun kohteena olevaan aiheeseen liittyen.

Tutkimuksessa pyritään vastaamaan muun muassa seuraaviin kysymyksiin: 1) Voi- daanko markkinasentimenttiä mitata tilastollisin menetelmin?, 2) Millainen aineisto on tarkastelun kannalta relevanttia: pörssitiedotteet, taloustuutiset, blogikirjoitukset, vai foorumit?, 3) Miten uutisten ja sentimentin muutokset näkyvät osakemarkkinoilla: koko markkinoiden tasolla, toimialoilla, yksittäisissä osakkeissa?, ja 4) Miten uutisvir- ran käsittelyssä voidaan hyödyntää ohjelmallisia työkaluja?

Lähteet: Malo, P., Sinha, A., Takala, P., Korhonen, P. and Wallenius, J. (2014): "Good Debt or Bad Debt: Detecting Semantic Orientations in Economic Texts." *Journal of the American Society for Information Science and Technology*. Vol. 65, No. 4, 782-796.

<http://onlinelibrary.wiley.com/doi/10.1002/asi.23062/abstract>

Malo, P., Sinha, A., Takala, P., Ahlgren, O., Lappalainen, I. (2013): "Learning the Roles of Directional Ex- pressions and Domain Concepts in Financial News Analysis." In: *Proceedings of IEEE International Confer- ence on Data Mining Workshops*, 2013. <http://sentic.net/sentire/2013/takala.pdf>.

Menetelmäkehityksen lisäksi tärkeää on muistaa muun muassa monitieteisyys niiden hyödyntämisessä. Moniin yhteiskunnallisiin ongelmiin voidaan hakea vastauksia uusien menetelmien avulla. Tämä vaatii eri tieteenalat ylittävää työskentelyä, muun muassa matemaatikkojen, ohjelmoijien ja muiden tieteenalojen edustajien kesken. Erilaiset lai- tos- ja tieteenalarajat tai näiden toiminta- ja metodologiaperinteet eivät saa estää uusien menetelmien sekä mahdollisuuksien täysimääräistä käyttöä. Yliopistojen tulisi olla valmiita muun muassa erilaisten monialaisten tutkimustiimien tukemiseen.

Erytyisesti kansainvälisessä yhteistyössä tehtävään ilmastonmuutoksen tutkimuksen ja ilmastopolitiikan toteuttamisen kannalta big data -menetelmät ovat oleellisessa roolissa. Big datan tärkeä merkitys on tunnistettu muun muassa YK:n ilmastonmuutokseen liitty- vissä hankkeissa.<sup>45</sup> Data sää- ja ilmastoherkistä toiminnoista mahdollistavat erilaisten, mahdollisesti aiemmin tuntemattomien yhteyksien löytämisen. Samalla tavalla saattaa

<sup>44</sup> Starr (2014), We Feel builds a real-time map of the world's emotions. <http://www.cnet.com/news/we-feel-builds-a-real-time-world-map-of-our-emotions/>

<sup>45</sup> United Nations Global Pulse: Big Data Climate Challenge. <http://unglobalpulse.org/big-data-climate/>; Bertolucci (2014), UN Unveils Big Data Climate Change Challenge. <http://www.informationweek.com/big-data/big-data-analytics/un-unveils-big-data-climate-change-challenge/d/d-id/1269545>

sääinformaation ja big data -analyysien yhdistäminen tuoda uusia keinoja informaation käytön tehokkuuden lisäämiseen.

### 3.4 Liikenne

Liikenteen ratkaisuilla tavoitellaan muun muassa liikenneturvallisuuden parantumista ja liikenteen tehostumista sekä liikenteen ekologisuutta. Ratkaisut voivat liittyä liikenteen ohjauksjärjestelmiin, logistiikan optimointiin sekä erilaisiin reittipalveluihin ja

Suomessa tehdään aktiivisesti liikennealan tutkimusta ja Suomi on kansainvälisesti kärkijoukossa esimerkiksi avoimen liikennedatan käytössä. Suomessa on tehty tutkimusta uudenlaisista liikennejärjestelmistä.

Älyliikenteen palvelutuotannossa tarvittavaa raakadataa tuottavat kuluttajat, yritykset ja julkinen hallinto. Dataa varastoidaan erilaisiin tietokantoihin. Digitaaliseen tietoon perustuvien liikenteen palvelujen ja sovellusten on arvioitu olevan maailmanlaajuisesti yksi tulevan vuosikymmenen kiinnostavimmista markkinoista. Suomella on mahdollisuus profiloitua liikenteen kokeilu-ympäristönä<sup>46</sup>. Tätä puoltaa muun muassa Suomen suhteellisen pienimuotoinen liikenne, jolloin pienempiä kokeiluja voidaan hyvin tehdä, usein myös normaalin liikenteen seassa.

Teknologisista trendeistä liikenteen kannalta ovat merkittävimmät ajoneuvojen kytkeytyminen tietoverkkoihin, mobiilien päätelaitteiden ja sovellusten määrän voimakas kasvu, jatkuva yhteys tietoliikenneverkkoon, tietovarantojen avautuminen, paikkatieto- ja navigointipalveluiden lisääntyminen, kohdennettu älykäs sisältö sekä päätelaiteriippumattomuus.

Oikea-aikaisten liikenteenhallinnan toimenpiteiden edellytys on laadukas ja kattava tilannekuva siitä, mitä liikenteessä juuri nyt on tapahtumassa, tai jopa ennakoiden, mitä on tapahtumaisillaan. Tilannekuva on koko älykkään liikennejärjestelmän perusta, jonka päälle voidaan rakentaa systeeminen kokonaisuus liikenteen ja liikkumisen ennakoivaan ohjaukseen. Liikenteen tilannekuvan laatu on suoraan riippuvainen kerättävän tiedon laadusta. Paikannusteknologioiden voimakas kehittyminen ja yleistymisen ovat tuoneet erilaiset paikantavat laitteet kaikkien liikkujien ulottuville. Autojen navigointilaitteet ja liikkujien omat mobiililaitteet mahdollistavat jo nykyisin eri liikkujaryhmien toimimisen tilannekuvan raakadatan tuottajina, ja tulevaisuudessa liikkujien tuottaman tiedon merkitys tulee kasvamaan varsin voimakkaasti. Käyttäjälähtöisen tiedonkeruun yhtenä keskeisenä etuna on kustannustehokkuus koko elinkaaren näkökulmasta, koska se ei vaadi investointeja kiinteisiin tienvarsirakenteisiin eikä myöskään huoltoon tai ylläpitoon.

Lähivuosina teknologiakehityksen myötä korostuu liikennejärjestelmän osien integroituminen toisiinsa, kun väylärakenteet ja ajoneuvot älyllistyvät ja liikkujien rooli tiedon tuottajana vahvistuu. Teknologioiden myötävaikutuksesta kehittyvä palvelusektori hyödyntää teknologiset innovaatiot ja tuottaa käyttäjien tarpeista lähteviä uusia sovelluksia ja palveluja. Kaikkialla ja koko ajan läsnä oleva tieto palveluineen mahdollistaa katkeamattoman yhteydenpidon ympäröivään maailmaan ja muihin liikkujiin, sekä luo mahdollisuuk-

<sup>46</sup> Liikenne- ja viestintäministeriö. Älyliikenne. <http://www.lvm.fi/alyliikenne>

sia uusille toimijoille ja liiketoiminnalle. Kun liikenteeseen ja logistiikkaan liittyvään tietoon saadaan yhdistettyä muista eri tietovarannoista saatava monipuolinen tieto, voidaan tietoja yhdistelemällä saada aikaiseksi uudenlaista tehokkuutta ja tuottavuutta sekä erityisesti kokonaan uusia innovatiivisia palveluketjuja.

Parhaimmillaan erilaiset liikenteeseen liittyvät big data -innovaatiot, kuten esimerkiksi liikenneanalyysi ja –ennusteet, voivat olla merkittäviä vientituotteita. Tutkimus- ja innovaatiotukea sekä muuta rahoitusta on liikenteen alalla saatavissa suhteellisen runsaasti. Näin ollen mahdollisuus kansainvälisestäkin merkittävään toimintaan on olemassa.

### **Älypuhelimet voisivat jo nyt tuottaa kattavaa liikennetietoa**

Liikenneverkon kattava reaaliaikainen käytön seuranta mahdollistaisi nykyistä tehokkaamman kapasiteetin hallinnan sekä käyttäjien informoinnin liikennetilanteesta. Hyvin informoidut liikkujat ja kuljetusten järjestäjät edesauttavat liikenteen sujuvuutta ja vähentävät liikenteen ympäristökuormitusta.

Myös siirtyminen yksityisautoista muihin kulkumuotoihin tulee houkuttelevammaksi, kun tarjolla on nykyistä suurempi määrä tietoa ympäröivästä liikennetilanteesta. Tällaista merkittävää tietoa syntyy jo nyt jokaisessa älypuhelimessa (määrä, nopeus, kiihtyvyys, jne.) ja tämän tiedon perusteella voitaisiin tuottaa kattavaa näkymää liikennetilanteesta niin valtakunnallisella kuin alueellisellakin tasolla. Edellytyksenä kuitenkin on, että käyttäjät suostuisivat jakamaan tällaista tietoa ja halutessaan määrittelemään sen sellaiseksi, että yksilön anonymiteetti säilyy.

Monille yritykselle tällaista tietoa kertyy jo nykyisin ja sen saaminen avoimesti yhteiskunnan käyttöön olisi merkittävä askel eteenpäin. Liikennetilanteita pystyttäisiin suurten datamäärien ansiosta analysoimaan ja sitä kautta tunnistamaan ongelmakohtia sekä ennakoimaan lyhyen aikavälin kehitystä. Tilastollisten menetelmien avulla tarkasteluun voisi seuloa eri kulkumuotoja (henkilöauto, bussi, raitiovaunu, pyöräily, jalankulku, jne.) ja näin kohdentaa tiedostusta ja palveluja hyvinkin tarkkaan segmentoidulle ryhmälle. Julkisen ja yksityisen sektorin hyvä ja avoin yhteistyö olisi välttämätön edellytys tällaisten tietomäärien saattamisessa kaikkia hyödyttävään käyttöön.

### **3.5 Teollinen internet**

Ennakoivat huolto- ja etäpalvelut ovat yksi sovellus laajemmassa kokonaisuudessa, josta usein käytetään termiä teollinen Internet (TI)<sup>47</sup>. Tällä tarkoitetaan sulautettujen ja älykäden 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.

Teollisen internetin sovellusalueita ovat muun muassa valmistavan teollisuuden prosessit ja niiden optimointi, energian käytön hallinta, käyttöomaisuuden hallinta ja ennakoiva huolto. Vähintään yhtä suuret hyödyntämismahdollisuudet ovat myös esimerkiksi tervey-

<sup>47</sup> Katso esim. VTT:n professori Heikki Ailiston esitys ja kirjoitus TI:stä VTT Blogissa. [https://www.tem.fi/files/39176/Heikki\\_Ailisto.pdf](https://www.tem.fi/files/39176/Heikki_Ailisto.pdf); <http://vtblog.com/2014/05/05/teollinen-internet-hype-vai-vallankumous/>

denhuollossa, kaupan ja logistiikan alueella, rakentamisessa ja kiinteistöjen hoidossa sekä kunnallisten ja muiden julkisten yhteiskuntapalvelujen tuottamisessa.

Yhteistä kaikille sovellusalueille on tavalla tai toisella automatisoitu linkki fyysisen maailman ja digitaalisen maailman välillä. Usein linkki syntyy joukosta antureita tai tunnisteita, 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 uusia palveluja ja liiketoimintaa.

### **Teollinen internet käytössä**

Monialayritys General Electricin valmistamista lentokoneturbiineista kerätään päivässä huimia datamääriä. Yksi turbiinin anturi kerää päivässä noin 500 gigatavua dataa. Yhdessä turbiinissa on 20 anturia ja GE:llä on noin 12 000 turbiinia.

Lähde: <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. Konecranes hyödyntää laajalti työkoneidensa tuottamaa anturitietoa, jota kerätään Konecranesin etäkeskukseen tallennusta ja analysointia varten.

Käyttäjillä on pääsy tietoihin huoltoaikataulujen optimoimiseksi ja turvallisuuden lisäämiseksi. Maailmanlaajuisen huoltoverkon kohdalla huoltokustannukset voivat koi-  
 tua kaukaisilla alueilla, kuten vaikkapa Afrikassa tai Australiassa, varsin mittaviksi. Anturit välittävät jatkuvasti tietoa osista, käytöstä ja huollon tarpeesta 2G-verkkoa hyödyntäen, joka on tarjolla koko maailmassa. Toiminnan seuraaminen ja tietomassojen kerääminen, varastointi ja analyysi pitää tuotteet jatkuvassa käytössä ja asiakkaat tyytyväisenä.

Lähde: <http://www.konecranes.com/resources/way-up-magazine/1-2012/from-smart-running-to-smart-maintenance> , [www.konecranes.com](http://www.konecranes.com), <http://www.bigdata.fi/artikkelit/big-data-ennakoiva-huolto>

Onnistuneesti kaupallistetuissa ja skaalautuvissa teollisen internetin palveluinnovaatioissa on myös huomattava vientipotentiaali. Suomella olisi mahdollisuuksia muun muassa palveluliiketoiminnan synnyttämisessä teollisen internetin avustuksella ja ympärille, ja teollisen internetin mahdollistamia uusia palveluita innovoimalla pienelläkin maalla on mahdollisuudet erottua.

Merkittävä osa tuotantoprosessiin ongelmanratkaisuista ja korjauksista voidaan nykyään automatisoida. Kaikki vähänkin arvokkaammat laitteet pystyvät itse analysointiin ja voidaan testata, analysoida, korjata etäältä. Usein vikaantuminen voidaan jo ennakoita. Tällä toiminnan optimoinnilla on mahdollista saada merkittäviä säästöjä. Etähuoltoratkaisujen avulla voidaan pitää ja saada lisää korkean jalostusarvon työpaikkoja Suomeen.

Teollisen internetin kehityksen hyödyntäminen vaatii osaamista ja yhdistelytaitoja erilaisilta alueilta. Näihin lukeutuvat muun muassa anturien kehitys, etäyhteyksien parantuminen sekä big data ana-

*”Teollinen internet ei ole mahdollisuus, se on pakko”*

lytiikka. Suomessa on alaa silmällä pitäen hyvä koulutustaso ja kompetenssi, vaikkakin osaajapula voi joidenkin alojen tai erityistaitojen kohdalla uhata tulevaisuudessa.

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. Esimerkiksi työ- ja elinkeinoministeriön viimeaikaisissa teollisuuspolitiikan ohjelmissa ja strategioissa teollinen internet on vahvana painopistealueena.

### 3.6 Puhdas teknologia

Suurten tietoaineistojen hyödyntäminen yhdistettynä puhtaan teknologian (Cleantech) osaamiseemme voi tarjota suomalaisille innovaatioille uusia vientimahdollisuuksia mm. Aasian suurille ja kasvaville vihreän teknologian markkinoille. Erilaisten puhtaan teknologioiden ratkaisujen kirjo on varsin laaja, mutta niiden ratkaisemat haasteet liittyvät varsin usein energian ja sen tuottamiseen vaadittavien resurssien optimointiin.

Ratkaisut ovat usein varsin dataintensiivisiä ja siten erittäin luonnollinen sovelluskenttä big data -teknologioiden hyödyntämiselle<sup>48</sup>. Puhtaan teknologian merkitys on yleisesti vahvassa kasvussa resurssiniukassa maailmassa. Puhdasta teknologiaa ei tule kuitenkaan ajatella niinkään omana ympäristötekniikan alanaan, vaan kaikkeen muuhun liiketoimintaan ja esimerkiksi tulevaisuudessa teollisen internetin palveluihin sulautettuna. Puhtaan teknologian kehitys kulkee pitkälti kytköksissä myös edellä kuvattuun teollisen internetin kehitykseen.

Puhtaan teknologian panostuksiin nähden Suomessa on suhteellisen vähän kansainvälisesti tunnistettuja, laajoja käyttöreferenssejä. Ratkaisuja täytyisi siis kokeilla rohkeasti ensin myös kotimaassa.

Puhtaan teknologian kehityksessä, kuten muillakin innovatiivisilla uusilla aloilla, suomalainen yrityskehitys tarvitsee välisiä verkottuvia yhteistyömalleja, joilla pienemmätkin yritykset voisivat toimia suoraviivaisemmin isojen markkinoilla jo olevien teollisuusyritysten innovaatiopartnereina.

---

<sup>48</sup> Sweet (2014), Big data, for the environment. <http://online.wsj.com/news/articles/SB10001424052702303369904579425063672446836>



Mikkelissä on käytetty jatkuvaa etämonitorointia ympäristön seurannassa, muun muassa Saimaan vesistön tilan osalta. Hanketta toteuttavat muun muassa Mikkelin ammattikorkeakoulu ja Observis Oy.

Lähde: [http://www.mamk.fi/ulkoiset/hanke/index.cfm?action=1&Hanke\\_ID=234](http://www.mamk.fi/ulkoiset/hanke/index.cfm?action=1&Hanke_ID=234)

Tanskassa Kööpenhaminan kaupunki ja Kööpenhaminan Cleantech-alan klusteri järjesti kilpailun, jonka avulla haettiin toteuttajaa big data -infrastruktuurille ja datan markkinapaikalle.

Lähde:

<http://www.cphcleantech.com/home/strategic-initiatives/ongoing-projects/big-data-infrastructure>

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

Lähde: [https://www.tem.fi/files/39772/VNP\\_kasvun\\_karjet\\_cleantech\\_ja\\_biotalous\\_08052014.pdf](https://www.tem.fi/files/39772/VNP_kasvun_karjet_cleantech_ja_biotalous_08052014.pdf)

### 3.7 Digitaalinen markkinointi ja digitaaliset kuluttajayritykset

Dataa ja analytiikkaa hyödyntävien yritysten liiketoimintatulosten on useissa tutkimuksissa havaittu olevan parempia kuin muiden<sup>49</sup>. Esimerkiksi Google, Amazon, Facebook ja LinkedIn ovat vuosien ajan systemaattisesti kehittäneet omaa big data -kyvykkyyttään ja hyödyntävät sitä nyt menestyksekkäästi sekä palveluidensa kehittämisessä ja optimoinnissa että relevanttien, kohdennettujen mainosten ja markkinointiviestien tarjoamisessa. Perinteiset toimijat eivät useinkaan ole ottaneet digitaalisia haastajia vakavasti, kunnes ovat yhtäkkiä löytäneet itsensä ahtaalta käyttäjien siirryttyä hyödyntämään uusia tuotteita ja palveluita.

Yhteistä menestyksekkäille digitaalisille kuluttajayrityksille on se, että liikkeelle on lähdetty tavoitellen ensisijaisesti kuluttajien sitouttamista ensiluokkaisella tuotteella ja kuluttajakokemuksella. Liiketoimintamallit ovat usein tulleet vasta myöhemmin. Hyvä kuluttajakokemus vetää puoleensa käyttäjiä ja investointirahaa. Dataa käytetään kuluttajakokemuksen jatkuvaan parantamiseen ja sisällön personointiin. Amazon haastoi perinteiset kirjakaupat Yhdysvalloissa, Netflix mullisti kuluttajien tavan katsoa TV:tä ja Spotify tavan kuunnella musiikkia. Viime vuonna Amazonin myynnistä 35 % tuli suosittelualgoritmien kautta.<sup>50</sup>

Netflixin *House of Cards*, 100 miljoonan Yhdysvaltain dollarin tuotanto, perustui kuluttajien digitaalisen katsomiskäyttäytymisen tarkkaan analyysiin. Zalandoilla yli 60 datatieteilijää hyödyntää Big Dataa käyttökokemuksen optimointiin, hinnoitteluun, myyntiin ja markkinointiin sekä jakeluketjun parantamiseen. Data- ja analytiikkakyvykkyudet ovat monille sijoittajille nykyään merkittäviä arvomuodostuskriteereitä, sillä ne lisäävät yrityksen tulevaisuuden arvoa.

<sup>49</sup> McCarthy, Rich & Harris (2011), signature clear space. Getting Serious About Analytics: Better Insights, Better Decisions, Better Outcomes. [http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture\\_Getting\\_Serious\\_About\\_Analytics.pdf](http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture_Getting_Serious_About_Analytics.pdf)

<sup>50</sup> Marshall (2006), Aggregate Knowledge raises \$5M from Kleiner, on a roll. <http://venturebeat.com/2006/12/10/aggregate-knowledge-raises-5m-from-kleiner-on-a-roll/>

Markkinoinnin lisäksi myös koko mediatoimialalla big datan hyödyntämisellä on suuri potentiaali, kuten alan yhden pioneeriyrityksen Netflixin esimerkki osoittaa. Mediatoimijat siirtyvät käyttämään erilaisia massadatan perustuvia menetelmiä myös sisältöjen kehittämiseksi. Lisäksi mediatoimialalla on tärkeä rooli big datan tuottajina.

Käyttäjäkokemuksen optimoinnista big datan avulla Suomesta löytyy huippuluokan osaamista. Rovion, Supercellin ja muiden edelläkävijäyritysten johdolla peliala on kasvanut huikeata tahtia<sup>51</sup>. Menestyksen yhtenä salaisuutena on reaaliaikaisen analytiikan oivaltava käyttö<sup>52</sup>, joilla pelin ladanneiden käyttäjien mielenkiinto saadaan pidettyä yllä. Pelaajien ja pelien käytöstä kerättävän tiedon (mm. sijainti, peliajat, saavutetut pisteet)<sup>53</sup> avulla käyttäjäkokemuksen parantaminen palvelun optimointi onnistuu selkeästi tehokkaammin. Vapaasti pelattavien (free-to-play), tuoton esimerkiksi pelissä etenemistä avittavien virtuaalisten hyödykkeiden myynnistä saavien pelien täytyy olla koukuttavia. Tarkka, varsinkin jatkuvasti peliä pelaavien käyttäjien tuntemus ja toiminnan ennustava analyysi on tällöin avainasemassa<sup>54</sup>.

Markkinointi on muuttumassa teknologiavetoiseksi. Gartnerin mukaan yritysten markkinoinnista vastaavat päälliköt käyttävät vuonna 2017 enemmän rahaa tietotekniikkaan kuin tietotekniikasta vastaavat<sup>55</sup>. Esimerkiksi LinkedInin ja Facebookin liiketoiminta perustuu isoihin käyttäjämassoihin ja kohdennettuihin ilmoituksiin ja mainoksiin. Kohdennetun mainonnan tehokkuus on huomattavasti parempaa verrattuna kohdentamattomaan. Retargetointi eli se, että ihmisille näytetään mainoksia tuotteista, joita kohtaan he ovat osoittaneet aiempaa kiinnostusta, esimerkiksi sähköisessä kaupassa, nostaa mainonnan tehokkuutta merkittävästi verrattuna siihen, että heille näytetään satunnaisesti valittuja tuotteita. Big datan tyypillisimmät käyttötapaukset markkinoinnin ja mainonnan alueella liittyvät kuluttajakokemuksen parantamiseen, sisällön personointiin ja viestinnän kohdentamiseen.

Yritysten markkinointipäälliköt etsivät jatkuvasti parhaita tapoja tavoittaa uusia asiakkaita ja myydä lisää nykyisille asiakkailleen. Tätä varten yritykset keräävät ja yhdistävät dataa eri lähteistä muodostaakseen mahdollisimman täydellisen kuvan asiakkaistaan. Toisin kuin perinteiset markkinointikanavat digitaaliset kanavat tarjoavat mahdollisuuden tehdä yksilöllisesti kohdennettuja markkinointitoimenpiteitä. Yritykset yrittävät tavoittaa tärkeimmät asiakasryhmänsä omissa kanavissaan vaikkapa suoramarkkinoinnin ja verkkosivujensa avulla sekä toisaalta julkaisijoiden kuten mediayritysten, hakukoneiden ja sosiaalisen median alustoilla. Julkaisijat puolestaan tekevät parhaansa tarjotakseen asiakasyrityksilleen mahdollisuuden kohdentaa mainontaa eri digitaalisissa ja mahdollisesti myös muissa kanavissa profiloitujen käyttäjiä käyttäytymisperusteisesti ja kontekstuaalisesti, muun muassa hakukone-, sisältöympäristöön kuuluvan natiivi-, ja sijaintiperusteisen mainonnan avulla.

<sup>51</sup> Tekes, The Game Industry of Finland. <https://www.slideshare.net/Tekesslide/finnish-game-industry>

<sup>52</sup> Kts. esim. Harris (2011), Rovio wants to predict Angry Birds players' moves. <http://gigaom.com/2011/08/19/rovio-trying-to-predict-angry-birds-players-next-moves/>; HP (2014), Supercell case study. [http://www.vertica.com/wp-content/uploads/2014/03/SuperCell\\_Case\\_Study.pdf](http://www.vertica.com/wp-content/uploads/2014/03/SuperCell_Case_Study.pdf)

<sup>53</sup> Kts. esim. Rovio.com: Analytics and data usage in Rovio games. <http://www.rovio.com/en/news/blog/235/analytics-and-data-usage-in-rovio-games>

<sup>54</sup> Dell: Tech Page One (2014), Big data behind big games. [http://techpageone.dell.com/technology/big-data-behind-big-games/#.U7FJAz1\\_t8E](http://techpageone.dell.com/technology/big-data-behind-big-games/#.U7FJAz1_t8E)

<sup>55</sup> <http://www.forbes.com/sites/lisaarthur/2012/02/08/five-years-from-now-cmos-will-spend-more-on-it-than-cios-do/>

Kohdennetun mainonnan ympärille on syntynyt myös erilaisia uusia toimijoita alkaen yrityksistä, jotka myyvät julkaisijoiden inventaariota mainospörssiin ja yrityksiin, jotka hallinnoivat mainostajien inventaariota. Kuluttajan kannalta on tärkeää, että toimialan oma sisäinen valvonta ja lainsäädäntö määrittävät läpinäkyvästi sen, mitä dataa kuluttajasta kerätään, miten ja kuka dataa käyttää. Myös yrityksille pelisäännöt ovat tärkeitä, sillä data on arvokasta eikä sen haluta valuvan vahingossa kolmansille osapuolille.<sup>56</sup>

Käyttäjien muuttuessa diginatiiveista datanatiiveiksi he odottavat tulewansa yksilöllisesti kohdelluiksi kaikissa kohtaamisissaan yrityksen kanssa. He odottavat yrityksen pystyvän tarjoamaan datan avulla samanlaista palvelua kuin ihminen, joka tuntee heidät henkilökohtaisesti. Ennustavaa analytiikkaa on joskus verrattu kyläkauppaan. Kyläkaupassa tuttu kauppias tietää, mitä asiakas yleensä ostaa ja laittaa ostoskorin valmiiksi. Kyläkauppias myös tietää, mitkä uudet tuotteet saattavat miellyttää asiakasta ja suosittelee niitä. Toisaalta asiakkaan odotukset ovat suuret. Kyläkauppiiaan tarjotessa jatkuvasti tuotteita, joista asiakas ei pidä, saattaa asiakas alkaa suosia lähikaupungin markettia.

#### **Opas big data-markkinointiin ja mainontaan**

Digimarkkinoinnin toimijoita yhteen keräävän IAB Finlandin Big Data Task Force julkaisi hiljattain uuden oppaan. "Datan hyödyntäminen digimainonnan kohdentamisessa" opastaa massadatan hyödyntämisessä. Oppaaseen tehdyssä selvityksessä huomattiin myös, että suomalaisten mainostajien, medioiden sekä toimistojen tämänhetkinen tietotaso mainonnan datalähtöisestä kohdentamisesta on kohtuullisen hyvä.

Lähde:

[http://www.iab.fi/media/pdf-tiedostot/standardit-ja-oppaat/iab-finland-datan-hyodyntaminen-6\\_2014.pdf](http://www.iab.fi/media/pdf-tiedostot/standardit-ja-oppaat/iab-finland-datan-hyodyntaminen-6_2014.pdf)

## **4. Julkishallinto big datan sovellusalueena**

Big data ajattelutapana ja teknologiana antaa hallinnolle uudenlaisia näkökulmia, joilla se voi edistää tavoitteitaan ja parantaa suorituskykyään, mikä lisää samalla kansalaisten tyytyväisyyttä julkisiin palveluihin.

Julkishallinnon palveluiden kehittäminen ja tuottaminen perustuvat entistä enemmän tiedon hyödyntämiseen. Kuten kaupallisessa palvelutuotannossa myös julkishallinnossa tieto on alusta, jonka päälle rakennetaan asiakaslähtöisiä ja kustannustehokkaita palveluita, jotka perustuvat käyttäjien ja sidosryhmien tarpeisiin sekä odotuksiin. Asiakaslähtöisyyden ydinedellytys on, että ymmärretään kansalaisen tarpeet ja kohdennetaan yhteiskunnan resurssit julkissektorin tietokehittämisessä oikeisiin asioihin.

Julkishallinnon toiminnoissa jatkuvasti kasvavan tietomäärän keskellä on tärkeää hahmottaa tiedon mahdollisuudet kokonaisuutena ja hallita merkityksellistä tietoa, eli ymmärtää, mitä tietoa on olemassa ja mitä puuttuu. Käynnissä olevan digimurroksen kes-

<sup>56</sup> IAB Europe ja EASA (European Advertising Standards Alliance) julkistivat 12.10.2012 European Interactive Digital Advertising Alliancen (EDAA), joka vastaa koko Euroopan laajuisesta selainkäyttämiseen perustuvan verkkomainonnan kohdentamisen itsesääntelyohjelmasta. Lisätietoja löytyy IAB Finlandin sivuilta [www.iab.fi](http://www.iab.fi)

kellä data- ja asiakaslähtöisen hallinnon tulee kyetä ratkaisemaan arkipäiväisiä tiedonhallintaan liittyviä ongelmia: missä ja kenellä tietoa on, onko tieto käytettävissä muodossa, miten voin liittyä tietovirtaan ja ymmärränkö tiedon käyttötarkoitukseeni oikein?

Julkishallinnon tietovarantojen yhteiskäyttöisyys ja hyödynnettäväksi saattaminen ovat tärkeitä arvoja. Kokonaiskuva tietopääomasta, tiedon ympärille rakentuva kehitysyhteisö, tietokehittämisen koordinointi ja standardit järjestelmäriippumattomat menetelmät

*”Myös julkinen sektori tarvitsee innovaattoreita, jotka voisivat etsiä substanssikohtaisesti big datan hyödyntämiskohteita”*

liikuttaessa tietoa ovat esimerkkitekijöitä, joilla rakennetaan julkishallinnon kokonaisuutta big data -ajattelun mukaiseksi.

Datalähtoisempää julkishallintoa voidaan yleisesti tarkastella esimerkiksi kolmella seuraavalla osa-alueella: Datalähtöisen päätöksenteon ja jatkuvan organisaatiokehityksen tuominen osaksi julkishallinnon kulttuuria, suorituskyvyn parantaminen ja kansalaisten henkilökohtaisemmat palvelut. Lisäksi tulee tarkastella big datasta saatavia hyötyjä eri hallinnonaloille kohdistettavissa olevien yksittäisten käyttötapauksien hyötyjen ja vaikuttavuuden kautta. Tässä raportissa ei ole käsitelty kansalliseen turvallisuuteen liittyviä käyttötapauksia ja erilaisia big data -aspekteja.

#### 4.1 Tavoitteena data- ja asiakaslähtöisempi hallinto

Datalähtöinen organisaatiokulttuuri edellyttää, että julkishallinto pyrkii tietoisesti ja jatkuvasti omaksumaan maailmalla parhaiten toimiviksi todettuja käytäntöjä tiedolla johtamisessa, organisaatioiden kehittämisprosesseissa, organisaation strategiaan sidotussa suorituskykymittauksessa sekä dataan perustuvan päätöksenteossa. Datalähtöisessä julkishallinnossa organisaatiot määrittelevät omaa missiotaan ja kokonaisstrategiaansa tukevat, selkeästi ymmärrettävät ja läpinäkyvät datalähtöiset suorituskykymittarit, jotka ohjaavat päätöksentekoa kohti organisaation ydinmission ja kansalaisten palvelutulosten toteutumista.

Reaaliaikaisia, organisaation ulkoisia ja sisäisiä monimuotoisia tietolähteitä analysoivat johtamisjärjestelmät mahdollistavat päätöksenteon tavalla, jossa organisaatiossa tekeillä olevien päätösten hyötyjä, riskejä, kustannuksia ja pitkän tähtäimen vaikutuksia on mahdollista ymmärtää ja simuloida ennen päätösten varsinaista toimeenpanemista. Big dataan pohjautuva tiedolla johtaminen mahdollistaa entistä reaaliaikaisemmat, hienoja-koisemmat ja tulevaisuuteen katsovammat mittarit, joiden avulla organisaatio kykenee entistäkin tehokkaammin ennakoivasti tunnistamaan omaa strategista tavoitetilansa mahdollisesti uhkaavat riskitekijät ja toimimaan suoraan havaittujen riskien välttämiseksi ja minimoimiseksi. Mittareiden pohjana olevien oletusten, käsittelysääntöjen ja lopulta itse datan läpinäkyvyys kansalaisille toimii omalta omaltaan takeena valittavien mittareiden luotettavuudelle.

Palvelun lisäksi valtion on luontevaa mitata myös itseään. Yleisiä, koko Suomea koskevia taloudellisia ja yhteiskunnallisia korkean tason mittareita ylläpitää esim. Findikaattori.<sup>57</sup> Toimintatapojen muuttuessa myös yhteiskunta voi saavuttaa suuria hyötyjä. Asiakasläh-

<sup>57</sup> [www.findikaattori.fi](http://www.findikaattori.fi)

töisyys ja ennakoiva palvelujen tarjoaminen voi tehostaa toimintaa, karsia byrokratiaa ja päällekkäisiä tehtäviä. Eri viranomaisten tietoja yhdistelemällä ennakoivampi palveluntarjonta sekä tarkempi kohdistaminen on niin asiakkaan kuin viranomaisten intresseissä.

### Ranskan datalähtöinen hallintouudistus

Esimerkkinä julkishallinnon mission asiakaslähtöistämisestä voidaan pitää Ranskan massiivista finanssikriisin jälkeistä hallintouudistusta. Uudistuksella oli mittavat säästö-tavoitteet: vuoden 2009–2013 aikana 15 miljardia euroa ja leikkaustarve yli 150 000 julkisen sektorin tehtävää.

Kustannussäästöjen lisäksi tavoitteena oli palvelun laadun 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 tapahtumista esimerkiksi "uuden yrityksen perustaminen". Uudistuksen alussa kansalaiset ja yritykset arvioivat kunkin julkishallintoon suuntautuvan asiointiprosessin koettua monimutkaisuutta arvoilla "yksinkertainen", "OK", "monimutkainen", ja "erittäin monimutkainen".

Varustettuna näillä alustavilla kansalais- ja yritysnäkökulmilla julkisen asioinnin havaitusta monimutkaisuudesta, tuli kyseisiä asiointiprosesseja hallinnoivien yksiköiden uudeksi ja ainoaksi asiakaslähtöiseksi missioksi annettiin uudistaa koko toimintansa siten, että kansalaisten ja yritysten monimutkaiseksi kokemat "elämäntapahtumat" saadaan mahdollisimman nopeasti muutettua mahdollisimman yksinkertaiseksi. Uudistuksen valmistuttua kyselyn tulokset osoittivat julkisen asioinnin havaitun monimutkaisuuden pudonneen (eli koetun palvelun laadun parantuneen) 20 prosenttia kansalaisten tapauksessa ja 25 prosenttia yritysten tapauksessa.

Lähde: [http://www.mckinsey.com/insights/public\\_sector/transforming\\_government\\_in\\_france](http://www.mckinsey.com/insights/public_sector/transforming_government_in_france)

## 4.2 Suorituskykyä parhailla käytännöillä

Julkishallinnon toimijoiden datalähtöisen päätöksenteon tehostamisen ja suorituskyvyn optimoimisen lisäksi big datalla avulla voidaan tehostaa samankaltaisten organisaatioiden ja yksiköiden keskinäistä suorituskykyvertailua (organizational benchmarking).

Organisaatioiden suorituskykyvertailun yleisenä lähtökohtana on se perusolettamus, että samankaltaisia perusprosesseja, tehtäviä tai rakenteita sisältävillä organisaatioilla on yleensä samankaltaisia toiminnan kehittämisen haasteita. Näin ollen vertailtavat organisaatiot voivat saada toisiltaan parhaita käytäntöjä oman toimintansa tehostamiseen. Toisilta oppiminen on luontevaa, koska mikään yksittäinen organisaatio ei yleensä ole paras kaikilla mahdollisesti vertailtavissa olevilla osa-alueilla.

Parhaita käytäntöjä voidaan löytää muun muassa sisaryksiköstä samasta emorganisaatiosta tai hallinnonalalta, mutta aivan yhtä hyvin niitä voidaan löytää julkisista tai yksityisistä toimijoista toiselta puolelta maailmaa.

Suorituskykyvertailut tarjoavat muille organisaatioille ja kiinnostuneille kansalaisille läpinäkyvän suorituskykyinformaation lisäksi vertailun kohteena olleille yksiköille vahvan kannustimen ja mitattavan lähtökohdan kehittää omaa suorituskykyään edelleen.

Big datan tuoma arvonnisa organisaatioiden väliseen suorituskykyvertailuun toteutuu sen mahdollistamien hienojakoisempien datalähtöisten suorituskykymittarien ja seurannan kautta. Ne puolestaan mahdollistavat vastaavasti aikaisempaa tarkemman vertailtavuuden sekä siten parhaiden käytäntöjen omaksumisen.

#### **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 ja ennakkoivampi 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.

Lähde: 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](http://www.mckinsey.com/client_service/public_sector/expertise/public_finance)

Lopulta datalähtöisen kulttuurin voivat saada toimimaan ainoastaan julkishallinnossa työskentelevät ja työnsä merkityksellisyydestä motivoituneet ihmiset, jotka ymmärtävät oman työnsä perimmäisen palvelutarkoituksen ja ovat valmiita datalähtöisyyden edellyttämään asennemuutokseen. Ihmisten kouluttaminen, motivoiminen ja vastuuttaminen tähän uuteen organisaatiokulttuuriin on suuri haaste ja muutos, jota on ajatusmallina lähestyttävä kokonaisvaltaisesti pelkkien organisaatiomuutosten sijaan. Datalähtöisen muutoksen käynnistäjiksi ja suunnannäyttäjiksi tarvitaankin kokeneita ja vastuutettuja tiedolla johtamisen edelläkävijöitä, joilla on vahvan datalähtöisen näkemyksen lisäksi kyky motivoida ihmisiä ja organisaatioita pitkäjänteisesti.

### **4.3 Kansalaisille yksilöllisempää julkista palvelua**

Julkishallinnon tavoitteena on tarjota kustannustehokkaampia ja älykkäämpiä palveluita kansalaisille eli säästää nykyisten palveluiden kustannuksissa ja parantaa niiden vaikuttavuutta.

Keskeinen keino parantaa palvelun vaikuttavuutta ja sitä kautta tyytyväisyyttä palveluihin on segmentoida kansalaiset big datan tarjoamin analyttisin menetelmin ja siten tarjota heille henkilökohtaisempaa palvelua. Toinen palvelun laatua parantava tekijä on sijainnin yhdistäminen tarjottavaan palveluun. Reaaliaikaisesti päivittyvää sijaintitietoa muihin tietolähteisiin yhdistelevät, tyypillisesti karttapohjaiset palvelut ovat paikannusteknologian yleistyessä kasvava sovellusalue myös julkishallinnossa.

Perustuen älykkään datalähtöisen palvelun näkemään big data -kokonaiskuvaan kansalaisesta - millainen kansalainen on, missä hän on, mitä kanavaa hän käyttää, miten hän palvelussa käyttäytyy ja mitä hän ylipäätensä on parhaillaan palvelussa tekemässä - voi palvelu tarjota kansalaiselle toimintokehotteita siitä, mitä hänen kannattaisi tehdä seuraavaksi. Henkilökohtaisemmat, tilanteeseen paremmin sidoksissa olevat ja kansalaisten omaa päätöksentekoa helpottavat palvelut parantavat entisestään käyttäjäkokemusta ja sitä kautta lisäävät kansalaistyytyväisyyttä ja luottamusta.

### **Saksassa työvoimahallinto löytää big datalla töitä**

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

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

Lähde: McKinsey Interview 2013, Frank Jurgen Weise, Behind the German jobs miracle.  
[http://www.mckinsey.com/features/government\\_designed\\_for\\_new\\_times/table\\_of\\_contents](http://www.mckinsey.com/features/government_designed_for_new_times/table_of_contents)

Edellä mainittujen ominaisuuksien lisäksi laadukkaiden palveluiden tulee aina kaikin tavoin säästää kansalaisten asiointiin liittyvää aikaa. Esimerkiksi kansalaisille näkyviä hakemus- ja tiedonsyöttöprosesseja tulisi automatisoida byrokratian ja käsittelyvirheiden vähentämiseksi.

Riippumatta tahosta, joka julkiset digitaaliset palvelut toteuttaa, tulee palveluiden tarjota rajapinta helposti kerättävissä olevaan reaaliaikaiseen palautteeseen niiden käytöstä, niin kansalaisten antaman henkilökohtaisen tai yhteisöllisen palautteen muodossa kuin palvelun varsinaiseen käyttöön liittyvän käyttäjä- ja käyttäytymisdatan muodossa. Palvelun käytöstä kerättyä palautetta tulee voida jälkepäin analysoida niin palvelun käytettävyyden kuin asiointin kehittämisen kannalta. Julkisen sektorin palvelujen kehitys tulee luonnollisesti tehdä läpinäkyvästi pitäen samalla huolta kansalaisten yksityisyyden suojasta.

## Julkishallinnon mahdollisia big data-käyttötapauksia

Julkishallinnolla on suuria määriä dataa rakenteisessa teksti- ja numeromuodossa. Julkishallinnon tuottavuutta voidaan parantaa yksittäisillä ja kustannussäästöjä teknisillä big data -ratkaisulla, jotka ovat kohdistettavissa julkishallinnon eri hallinnonalojen ja yksiköiden täsmällisiin tietojenkäsittelyllisiin ongelmiin. Hyödynnettäessä dataa julkishallinnossa, on tietoteknisiä ratkaisuja huomattavasti tärkeämpää saada aikaan datalähtöisyyden edellyttämä tietoisuuden lisäys 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 kohdistamiseksi tarjolla oleviin ja työttömille hakijoille yksilöllisesti soveltuviin työ- ja jatkokouluttautumistilaisuuksiin. Tällä voidaan tukea hallituksen tavoitteita nuorisotyöttömyyden torjumisessa, työurien pidentämisessä, ammatillisen koulutuksen tarjoamisessa, välityömarkkinoissa sekä työtarjousten lisäämisessä
- 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
- Parhaiden käytäntöjen soveltaminen yhdessä yritysten kanssa työttömyyden suunnan ennakoimiseksi
- Työllisyyteen liittyvien koulutus- ja osaamistarpeiden tilannekuvaseuranta työvoimapolitiittisten strategioiden paremmaksi ennakoimiseksi

### Harmaan talouden ja kansainvälisen veronkierron torjunta

- Tehostettu petosten ja virheiden havainnointi big data -ratkaisulla osana harmaan talouden torjuntaohjelmaa
- Kovaan dataan perustuvien veroanalyysien täydentäminen esimerkiksi sosiaalisen median datalla

### Hallituksen kestävä kehityksen mittarit

- Kestäväan kehitykseen liittyvät big data -mittarit ja niiden seuranta
- Ympäristön kestävyteen liittyvät mittarit ja niiden kansalaisyhtävällinen 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 visualisoinnit avoimen datan pohjalta, esimerkiksi oman äänestyspäätöksen helpottamiseksi.

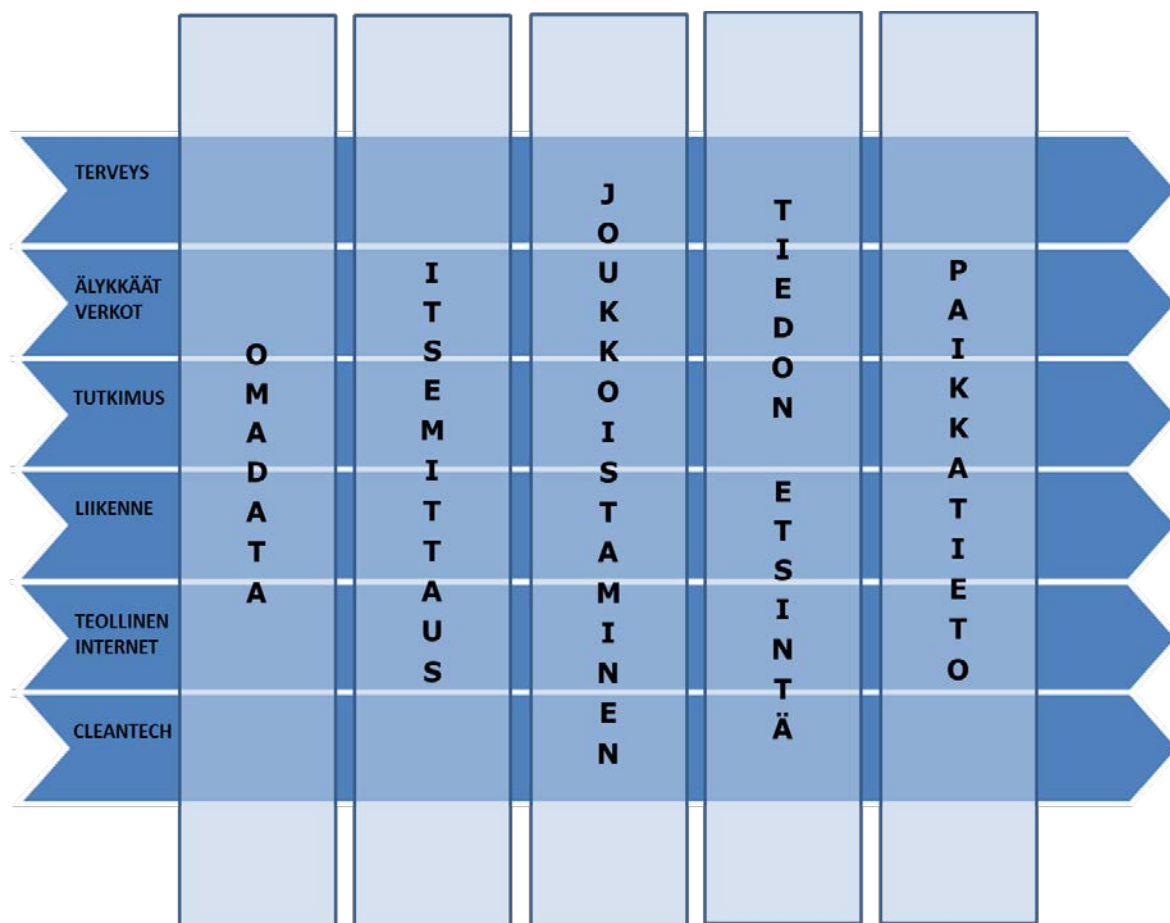


## 5. Sovellusalueet läpileikkaavia teemoja

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.

Läpileikkaavat teemat eivät suoraan edusta yksittäistä toimialaa, vaan lupaavia datalähtöisiä toimintamalleja, joita sovelletaan useilla toimialoilla. Läpileikkaavien alojen edistäminen on tärkeää myös siksi, että ne mahdollistavat usean eri sovellusalan kehittymistä. Läpileikkaavat teemat vaikuttavat kaikilla edellisessä luvussa tunnistetuilla sovellusalueilla. Teemat ja sovellusalueet on esitetty kuvassa 8.

**Kuva 8. Sovellusalueet ja läpileikkaavat teemat**



### 5.1 Omadata

Omadata (mydata/midata) viittaa toimintakulttuuriin, jossa organisaatio palauttaa keräämäänsä yksilökohtaista tietoa ihmiselle itselleen. Tämän jälkeen ihminen voi hyödyntää näin samaansa tietoa suoraan tai jakaa sen edelleen haluamallaan tavalla. Jaettu tieto voi olla analysoimatonta raakadataa tai se voidaan jalostaa esimerkiksi visualisoidulla muotoon, jossa ihminen itse ja muut voivat sitä hyödyntää. Kansalaisten tulee voida nykyistä paremmin vaikuttaa siihen, mihin ja miten heitä koskevia tietoja käytetään.

Hänen tulisi halutessaan voida hallita omia tietojaan ja voida luovuttaa kerättyä tietoa myös eteenpäin.

Oman tiedonhallinta kehityksen edellytyksenä nousee jatkossa yhä useammin esiin. Hiljattain esimerkiksi Yhdysvaltain liittovaltion kauppakomissio FTC suositteli kongressille, että kuluttajilla pitäisi olla enemmän kontrollia heistä kerättyihin tietoihin ja niiden käyttöön. FTC oli huolissaan kuluttajadataa keräävien ja myyvien yritysten nykyisistä toimintatavoista.<sup>58</sup>

Omadatan hyödyntäminen mahdollistaa uudenlaisten palveluiden kehittämisen. Erillisillä pyynnöillä tai analyyseillä kerättyjen tietojen laajempi, tarkempi ja kansalaisten itsensä hyväksymä tietoaiteistojen hyödyntäminen on myös tietoa keräävien yritysten intresseissä. Samalla täytyy kuitenkin pitää huoli siitä, että myös ilman tietojen luovuttamista yhtä laadukkaita palveluita on saatavilla. Tilanne, jossa laaja omien tietojen luovuttaminen on ylipäättään edellytys koko palvelun saamiselle, ei ole siten toivottava.

Niin sanotut omadata-ratkaisut voisivat 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 käyttäjän kannalta omien tietojen hallinta ja hyödyntäminen mahdollistaa muun muassa kustannus- ja aikasäästöjä, henkilökohtaisten palveluiden räätälöinnin. Lisäksi läpinäkyvyys tietojen keräämisessä ja kansalaisten oma hallinta lisäisi luottamusta tietoa kerääviä yrityksiä kohtaan.

Omadata-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. Läpinäkyvyys tietojen keräämisessä ja kansalaisten oma hallinta lisäisi luottamusta tietoa kerääviä yrityksiä kohtaan.

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

Omadata-aloitteisiin liittyvä keskustelu yksityisyydestä korostaa ihmisten valtaa päättää siitä, kuka heidän tietojaan saa käyttää sen sijaan, että ihmiset antavat kerralla suostumuksen kaikkien aineistojen käyttöön. Digitaalisuuteen liittyy nimenomaan mahdollisuus monistaa ja käyttää tietoa yhä uudelleen, jolloin ihminen ei suostumustaan antaessaan voi tietää tai hallita tulevaisuuden käyttötarkoituksia.

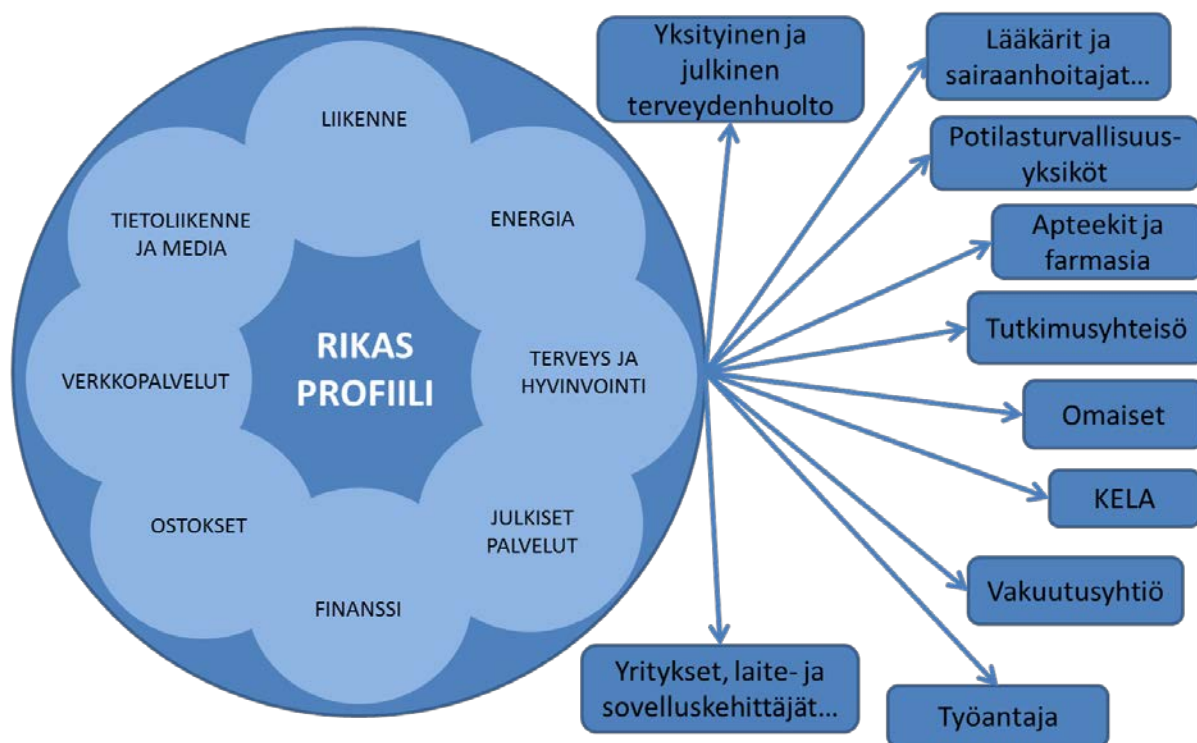
Toimivan omadata-toimintaympäristön ja –kulttuurin luominen edellyttää, että yhteisistä henkilökohtaisiin aineistoihin liittyvistä periaatteista pystytään sopimaan. Tarvitaan myös sääntelyä, jonka avulla aineistojen jakaminen on mahdollista. Tarvitaan koneluettavia

<sup>58</sup> Federal Trade Commission (2014), Data Brokers: A Call for Transparency and Accountability. <http://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf>

rajapintoja aineistojen jakamisen tueksi, standardeja ja palveluita tiedon hallittuun siirtämiseen, varastointiin, käsittelyyn ja analysointiin.

Oheisessa kuvassa 9 on esitetty omadatan käyttöön liittyviä näkökulmia. Omadatan ideana on se, että ihmisellä käyttäjänä olisi uudelleen hyödynnettävissä olevassa muodossa ja hallussaan itseään koskevaa tietoa sekä siten mahdollisuus antaa tietoja haluamalleen taholle. Yksilön kannalta omien tietojen, esimerkiksi tässä terveystietojen hallinta antaa paremman tietopohjan tehdä terveyttä koskevia valintoja ja siten tarvittaessa muuttaa käyttäytymistään. Yksilön suostumuksella annettujen tietojen käyttö sovelluskehityksessä mahdollistaisi puolestaan palveluinnovaatioita ja palvelujen rikastamista. Näin yksilö ei olisi myöskään sidottu yksittäiseen palveluntarjoajaan, sillä asiakastapahumiin liittyvät tiedot olisi siirrettävissä uudelle palveluntarjoajalle. Omadata-ratkaisut myös edistäisivät tiedon yhteiskäyttöisyyttä, sillä esimerkiksi viranomaiset ja erilaiset palveluntarjoajat voisivat kehittää ja yhdistellä helpommin käyttäjälähtöisiä palveluja datan ollessa yksittäiseen profiiliin yhdistettävissä.

**Kuva 9. Omadata, esimerkkinä terveys- ja hyvinvointiala**



Käyttäjän suostumuksella käyttäjistä voisi saada rikasta profiilitietoa, jonka vastineeksi voitaisiin tarjota käyttäjille parempia palveluita

Lähde: Mukailtu OKFN Finlandin ja Jyväskylän yliopiston kuvista

## 5.2 Itsensä mittaaminen

Itsensä mittaamisen (quantified self) keskeisenä ajatuksena on yksilöllisen mittaustiedon hyötykäyttö terveyden ja hyvinvoinnin edistämiseksi. Viime vuosien aikana itsensä mittaamisen trendi on levinnyt eri puolille maailmaa, ja saanut uusia muotoja. Kyse ei ole enää yksinomaan superterveydestä haaveilevien teknoilijoiden liikkestä vaan mittaa-

mista istutetaan yhä uusille elämänalueille. Big data ja itsensä mittaus kytkeytyvät läheisesti toisiinsa, koska itsensä mittauksesta syntyy big dataa.

Digitaalisten itsensämittausaineistojen arkistointiin tarvitaan uusia ratkaisuja. Tiedonhallinta on turvattu, kun yksilö omistaa tietonsa ja antaa sille käyttöoikeuden halutessaan. Myös digitaalialan suuryritykset ovat heränneet itsensämittaukseen ja siinä tarvittaviin ratkaisuihin.

Itsensä mittauksen ympärillä on paljon uusia aloitteita, joita motivoi myös terveydenhuollon toimialamuutos. Suomessa itsensä mittausteknologioita 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 esimerkiksi seurata unensa laatua. Kehityskulku voi vähitellen muuttaa terveydenhuollon painopistettä ja lääkärin työtä. Lääkärit voivat keskittyä erikoistuneeseen sairauksien hoitoon ja ennaltaehkäisyyn, kun terveyden ylläpitoon on uudenlaisia vaihtoehtoja.

Itsensämittaus haastaa myös muilla tavoin nykyisiä käsityksiä oman terveyden hallinnasta. Itsensä mittaajat ovat osoittaneet, että esimerkiksi ajankäyttödata voi olla terveystietoa. Esimerkiksi riippuvuuksia voi tarkastella ajankäytön avulla. Itsensä mittaajatus on leviämässä hyötykäyttöön myös ympäristön tilan arvioinnissa ja hoidossa. Voiko ilmanlaadusta ennakoita tulevaisuudessa keuhkohtaumapotilaan kohtauksen?

Itsensä mittaustietoa on alkanut levitä myös muille toimialoille, esimerkiksi oman talouden hallintaan tai koulutukseen. Mittatietoa voi käyttää esimerkiksi omaehtoisen oppimisen seurantaan tai tekemään näkyväksi opettajien työtä eri luokkien kanssa.

### **Itsensä mittausta myös pankkipuolella**

Osuuspankin Pivo-mobiilipalvelun avulla asiakas voi seurata omaa kulutustaan ja omia kanta-asiakkuuksiaan.

Lähde: <https://www.op.fi/op/henkiloasiakkaat/opastus/uusi-mobiilisovellus-pivo-on-vastaus-kaupankaynnin-muuttamiseen?cid=151725214&srcpl=3>

### **5.3 Joukkoistaminen**

Demokraattisen osallistumisen mahdollisuuksia sekä kansalaisten äänen näkymistä esimerkiksi päätöksenteossa voidaan lisätä big datan avulla. Teknologia ja uudet tiedon analysointimenetelmät tulisi valjastaa tukemaan myös demokratian ja kansalaislähtöisyyden kehitystä. Tämä olisi esimerkillistä ja avoimen yhteiskunnan toimintaa. Toisaalta kokemuksia kartuttamalla voitaisiin kehittää myös vientikelpoisia konsepteja.

Big data tarjoaa paljon mahdollisuuksia myös kansalaisten 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 voi saada paremman kuvan kansalaisten mahdollisista ongelmista ja esimerkiksi reaktioista erilaisiin päätöksiin. Menetelmien avulla voisi löytää piileviä aiheita ja tehdä päätöksenteosta ja hallinnosta mukautuvampaa sekä kansalaislähtoisempää.

Samalla joukkoistaminen on 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 antava liikennesovellus, joka yhdistelisi sekä käyttäjien dataa että esimerkiksi reaaliaikaista liikennedatata saavuttaakseen mahdollisimman tarkat ennusteet sekä paremman palvelun.

### **Tulevaisuuskatsauksen joukkoistus**

Liikenne- ja viestintäministeriön vuoden 2014 tulevaisuuskatsauksen tausta-aineiston tuottaminen toteutettiin tavanomaista sidosryhmäkuulemista laajemmin kaikille avointa fasilitoitua joukkoistustapaa hyödyntäen.

Tulevaisuuskatsauksen sisältöä pääsi siis kuka tahansa kirjoittamaan sekä kommentoimaan vuorovaikutteisesti avoimella wikialustalla, jossa työtä ohjasivat ministeriön virkamiehet ja jonne sidosryhmiä ohjattiin aktiivisesti eri sosiaalisen median kanavista. Koko aineisto jätettiin avonaiseksi wikiin, joten sen myöhempi hyödyntäminen avoimena datana on jatkuvasti mahdollista. Sen lisäksi, että tuotetut tekstiaineistot täyttävät avoimen datan kriteerit, päätettiin materiaalista tehdä myös linkitettyä dataa (linked open data).

Toisin sanoen eri toimenpide-ehdotukset liitettiin niin ikään joukkoistamalla toisiinsa. Liitosten lisääminen eli datan semantisointi ja rikastaminen avainsanoilla mahdollistaa sen, että laajojen tekstimassojen, kuten hallinnon toimenpideohjelmien keskinäistä yhteensopivuutta, mahdollisia sisäisiä ristiriitaisuuksia ja vaikutuksia, voidaan arvioida aiempaa syvällisemmin. Myös automatisoidun tekstidata-analyysin ja tiedon louhinnan menetelmien käyttäminen tulevat mahdollisiksi yllä kuvatussa toimintamallissa, jossa pelkkien pdf-raporttien sijaan yhdistellään hallinnon omissa prosesseissa syntyvästä datasta koneluettavaa big dataa.

## **5.4 Tiedon etsintä**

Tietomäärien valtava lisääntyminen luo päivittäisiin työtehtäviin lisähaasteita tietotyöntekijöille. Tiedon etsintä on alue, jota paremmin tukemalla voidaan saavuttaa merkittäviä säästöjä. Erilaisille tietotulvaa helpottaville työkaluille ja etsintämenetelmille on monikanallisten yritysten hallitsevasta asemasta huolimatta vielä kysyntää ja tilaa kehittämiselle esimerkiksi tutkimuksen parissa. Tietotyön lisääntyessä jatkuvasti erilaisiin tarpeisiin räätälöidyillä, tiedon etsintää helpottavilla työkaluilla on tarvetta monilla aloilla.

Suomessa tehdään myös merkittävää alaan liittyvää tutkimusta. Tiedon etsintään liittyvät myös semanttisen verkon mallit. Suomessa alan tutkimuksella ja teknisten ratkaisujen kehittämisellä on pitkä, vuosituhatvuotinen alkuun ulottuva historia.

Tiedon etsinnän alalta voi oikeilla panostuksilla löytyä hyvin merkittäviä 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.

### 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.

Lähteet:

[http://www.tiede.fi/artikkeli/uutiset/suomalaistutkijat\\_tekivat\\_uudenlaisen\\_hakukoneen](http://www.tiede.fi/artikkeli/uutiset/suomalaistutkijat_tekivat_uudenlaisen_hakukoneen);  
<http://research.ics.aalto.fi/mi/online-papers/Ruotsalo13cikm.pdf>

### FINTOsta löytyy eri aihealueiden ontologisia luokituksia

Valtiovarainministeriön ja opetus- ja kulttuuriministeriön yhteisessä ONKI-projektissa kehitetään FINTO-palvelua. FINTO on Kansalliskirjaston ylläpitämä, avoimia rajapintoja jatkohyödyntämiseen tarjoava alusta erilaisten sanastojen julkaisulle ja selaamiselle.

FINTO perustuu Aalto yliopiston Semanttisen laskennan tutkimusryhmän tuloksiin.

Lähteet: <http://finto.fi/fi/>;

<http://wiki.helsinki.fi/display/ONKI/ONKI-projekti+ja+suomalainen+ontologiapalvelu+Finto>

## 5.5 Paikkatieto

Paikkatieto on tietoa, jolle voidaan osoittaa sijainti. Paikkatieto koostuu kohteen tai ilmiön sijaintitiedosta tai ominaisuustiedosta. Viime vuosina paikkatiedon määrä on muun muassa sosiaalisen median, älypuhelimien ja laajalti yleistyneen GPS-tekniikan ansiosta lisääntynyt merkittävästi. Samaan aikaan paikkatietoa tuottavat laitteet, standardit ja analysointimenetelmät itsessään ovat kehittyneet tuottamaan ja jalostamaan entistäkin tarkempaa ja hyödynnettävämpää sijaintitietoa.

Tähän asti paikkatietoja on hyödynnetty erityisesti luonnonvarojen kartoittamisessa, maankäytön suunnittelussa ja ympäristönsuojelussa. Paikkatiedot kuitenkin tarjoavat lukemattomia sovellusmahdollisuuksia lähes kaikilla liiketoimintasektoreilla ja julkisen sektorin hankkeissa.

Paikkatietojen innovatiivinen käyttäminen tarjoaa aivan uusia näköaloja muun muassa kuljetuspalveluiden, yhdyskuntasuunnittelun ja liiketoiminnan suunnittelun suhteen. Palveluverkostojen optimaalinen kohtaaminen palveluiden käyttäjien kanssa voidaan suunnitella tarkemmin. Yksinkertaisimmillaan paikkatiedoilla voidaan kohdentaa mahdolliselle käyttäjälle tai asiakkaalle täsmällistä informaatiota tarjolla olevista lähipalveluista.

Paikkatiedoilla voidaan kiinnittää huomiota muun muassa ihmisten kulutustottumusten visualisoimiseen kartalla ja luoda näiden tietojen perusteella malleja joukkojen ominaisuuksista. Palveluyritys voisi 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 ennakoimisen apuvälineeksi. Paikkatietojen avulla voidaan myös tehostaa ja parantaa luonnonvarojen kartoitusta, maankäytön suunnittelua, ympäristön suojelua ja sääennusteita. Paikkasidonnaisten sää-tietojen lisääminen liiketoiminta- ja muuhun paikkatietoon mahdollistaa uusien innovatiivisten palveluiden syntymistä.

### **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 pilvipalvelun, 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.

Lähteet:

[http://www.esri.fi/referenssit/srv-helppokayttoinen\\_paikkatietojarjestelma\\_kiinteistoliiketoiminnan\\_tukena/](http://www.esri.fi/referenssit/srv-helppokayttoinen_paikkatietojarjestelma_kiinteistoliiketoiminnan_tukena/)

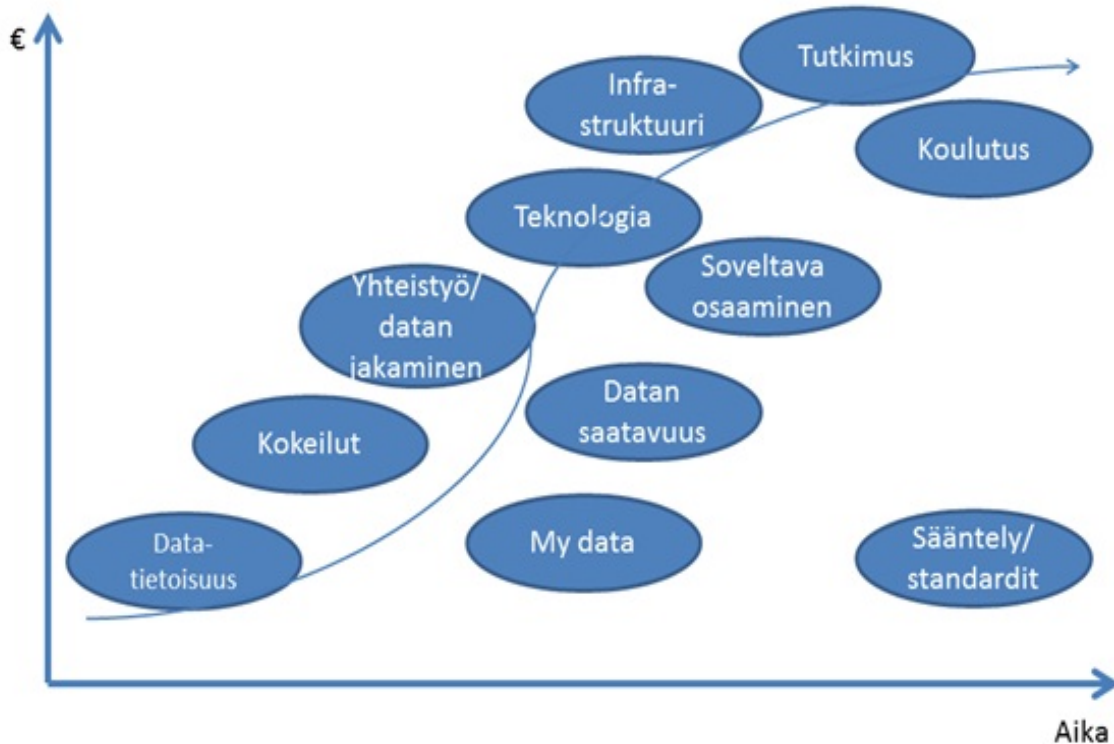
## **6. Toimenpiteet**

Tässä luvussa esitellään työryhmän ehdotuksia toimenpiteiksi, joita tarvitaan laajojen tietoaineistojen hyödyntämiseen eri sovellusalueilla (luvut 3-5) ja siinä tarvittavien edellytysten (luku 2) parantamiseksi kansallisesti.

Alla olevassa kuvassa on alustavasti arvioitu raportissa esitettyjen toimien vaikuttavuutta ajassa. Esimerkki kuvaa sitä, miten välittömästi toimet vaikuttavat: kokeilut vaikuttavat lyhyellä aikavälillä, koulutus varsin pitkällä ja niin edelleen. Toimien toteutusta voidaan tehdä nykyisten resurssien puitteissa, osa toimenpiteistä vaatii erillistä resurssointia. Toimeenpanossa on varmistettava, että saavutetaan vaikuttavuutta sekä lyhyellä että pidemmällä aikavälillä ja että toimenpiteet tukevat toisiaan. Toimenpiteiden priorisointi, vastuutus ja aikataulu määritetään toimeenpanosuunnitelmassa.



Kuva 10. Toimien vaikuttavuus ajassa



## 6.1 Datatietoisuus

Big datan hyödyntämistä tarvitaan kaikilla sektoreilla, 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 revolutiota. Erityisesti yrityksissä tarvitaan siten toimenpiteitä ymmärryksen lisäämiseksi liiketoiminnan muutoksesta dataintensiivisessä kilpailussa. Jokaisen organisaation päätöksenteon tulisi muuttua tietoon perustuvaksi (nk. data driven -pätöksenteko).

### Toimenpide-ehdotukset:

*”Datalähettiläitä” hankitaan ja käytetään tuomaan big data -tietoutta ja osaamista yrityksiin ja hallintoon muun muassa toteuttamalla kansainvälisiä vierailijaluentoja ja osaajavaihtoa.*

*Luodaan alan yrityksiä, tutkijoita ja asiantuntijoita kokoava avoin big data -klusteri sekä osaamisen kehittämistä tukevaa verkostotoimintaa tämän ympärille. Klusterin toimijat toimivat kansallisena vaikuttajaryhmänä alan kehityshankkeissa ja toiminnan valmistelussa.*

*Julkisen hallinnon organisaatioihin nimetään datavastaavia, joiden tehtäviin kuuluu datan keruun ja analysoinnin kehittäminen sekä hyvien käytäntöjen tunnistaminen ja levittäminen.*



## 6.2 Koulutus ja osaaminen

Korkeatasoinen osaaminen on olennaisin big dataa edistävä tekijä. Sen puute on toisaalta kehitystä rajoittava tekijä.

Koulutus on keskeinen väline uusien kasvualojen kehittämisessä. Ratkaisuna on toistaiseksi käytetty lähes yksinomaan tutkintoon johtavan koulutuksen kehittämistä korkeakouluissa ja toisella asteella. Täydennyskoulutusta on kehitettävä elinikäisen oppimisen periaatteen mukaisesti erityisesti tiedonkäsittelyn taitojen osalta, jossa kehitys on nopeaa ja työelämässä taidot myös vanhenevat.

Big datan hyödyntämisessä tarvitaan sekä teknistä että toimialaspesifistä osaamista. Datatieteilijältä odotettavat taidot (kuvailua edellytyksissä) ovat poikkitieteellisiä, joten tiimien osajapoolien rakentaminen voisi olla yhtenä ratkaisuna osajapulaan. Koulutuksessa monialaisuus tulee huomioida.

Data-analytiikka -osaamista on saatava siten kaikille koulutustasoille. Korkeakoulututkinnon yleisiin opintoihin tulisi saada datan hallinnan ja käsittelyn painotusta oppiaineesta riippumatta. Oppilaitosten ja korkeakoulujen tulisi paremmin erikoistua ja tehdä yhteistyötä esimerkiksi verkkomuotoisen opiskelun kehittämiseksi. EU:n tavoitteena on luoda Euroopassa yliopistojen verkosto datataitojen osalta, johon kansallista opetus- ja tutkimusyhteistyön kehittämistä on linkitettävä.

On kehitettävä julkisrahoitteisia ja kaupallisesti tarjottavia täydennyskoulutusohjelmia big data -asiantuntijoille. Yrityksissä osaamista voitaisiin kehittää mm. luomalla erilaisia harjoittelujaksoja ja projekteja alan opiskelijoille, tutkijoille ja työmarkkinoilla vapaana oleville osaajille.

### **Toimenpide-ehdotukset:**

***Kehitetään käytännön ratkaisuja ja sisältöjä siihen, että opettajankoulutukseen ja peruskoulun opetussuunnitelmien toteutukseen sisältyy vuonna 2016 dataan käyttöön liittyvää opetusta ja tuetaan avoimen lähdekoodin teknologioiden leviämistä.***

***OKM-ICT2015-ryhmä ja Opetushallitus kartoittavat big dataan liittyvää koulutustarjontaa.***

***Kannustetaan keskiasteen oppilaitoksia ja korkeakouluja luomaan tarvittavia osaamisprofilleja ja koulutusohjelmia (omia ja yhteisiä) big datan hyödyntämisen alueelle.***

***Osallistutaan EU-laajuisen big dataan erikoistuvien yliopistojen verkoston luomiseen.***

***Tuetaan big data -opetusympäristöjen rakentamista.***

***Ohjelmoinnin/tietojenkäsittelyn/liiketoimintaprosessien opiskelijoita kannustetaan yrityksiin big data -harjoitteluun, -projekteihin ja -opinnäytetöihin.***

***Täydennys- ja muuntokoulutusta uudistetaan vastaamaan big data -osaamistarpeisiin ICT2015 -työryhmän esittämällä tavoilla.***

***Tuetaan yritysten big data -osaamista kehittävien sisäisten hautomojen toimintaa, osaamisvaihtoa ja työnkiertoa.***

### **6.3 Tutkimus**

Suomessa big data -tutkimusalan huippuosaamista on useissa yliopistoissa ja tutkimusorganisaatioissa, mutta tutkimus on suhteellisen koordinoimatonta. Monialaisten big data -tutkimuskeskusten ja yhteistyöverkostojen syntyminen vauhdittaa alan kehitystä.

Tutkimuksen rahoituksessa tulee antaa myös enemmän painotusta hankkeen aikana tuotettavalle datalle ja sen laadulle. Yksittäisiä tutkimushankkeita tulee siten motivoida tuottamaan yleiskäyttöisempää dataa. Tutkimusten toteutuksissa tulee kannustaa yhteistoi-  
mintaan erityisesti suuria resursseja vaativissa hankkeissa. Tutkimusrahoitusta tulee kohdentaa myös avoimen tiedon keruun ja jakamisen kriteereiden mukaan. Esimerkiksi biopankkitoiminnassa materiaalien tehokkaan hyödyntämisen näkökulmasta tulisi mittaukset suorittaa siten, että niissä tuotettu data palvelisi mahdollisimman montaa tutkimushanketta.

Tutkimuksen kehittämisessä voitaisiin harkita myös tutkimushankkeiden palkitsemista siitä, jos muut tutkijat ovat sen ottaneet tuottaman datan hyötykäyttöön. Rahoituksen tulostavuuksiin voidaan kehittää mallia, jossa datan jaettua hyödyntämistä seurataan.

#### **Toimenpide-ehdotukset:**

***Ohjataan tutkimusrahoitusta erityisesti monitieteiseen ja -alaiseen big data -tutkimustoimintaan, huippututkimuskeskittyymiin sekä tutkimusta ja datan hyödyntämistä kehittäviin keskuksiin.***

***Huolehditaan suurien tietoaaineistojen käyttävän tutkimuksen tutkimusinfrastruktuurin riittävästä tulevaisuuden tarpeista huomioiden.***

***Parannetaan suurten tutkimustietoaaineistojen saatavuutta Avoimen tieteen ja tutkimuksen hanke (OKM) puitteissa. Hyödynnetään hankkeessa kehitettyjä palveluja ja käytäntöjä.***

***Kannustetaan rahoituksella ja ohjauksella eri toimijoita maksimaaliseen tutkimusdatan jakamiseen ja siten tutkimuksen tehostamiseen läpi tutkimushankkeiden elinkaaren***

***Lisätään resurssien puitteissa kansallista osallistumista tutkimuksen tietoaaineistojen jakamisen kansainväliseen kehitykseen ja standardointiin globaaleissa yhteistyöelimissä.***

## 6.4 Infrastrukturi

Big datan hyödyntäminen vaatii korkealuokkaista tieto- ja viestintäinfrastruktuuria. Perusedellytysten, kuten viestintä- ja tiedonsiirtoyhteyksien, pilvikapasiteetin ja datavarastoinnin ratkaisujen on oltava kansainvälistä huipputasoa. Infrastruktuuritarpeita on ennakoitava niin yksittäisten sektorien kuin yhteishankkeiden kannalta.

Big datan käsittelyä, jakamista ja hajautusta tukevien kansallisten infrastruktuurien kehittämistä on tuettava. Infrastruktuureihin liittyvällä kehitystyöllä voidaan tarkoittaa esimerkiksi:

- tiedonkäsittely- ja laskentainfrastruktuurien kehittämistä
- yhteisiä datan hallintajärjestelmiä ylläpitoon ja laadunvarmistukseen
- T&K -yhteistyön ja arvioinnin kehittämistä erilaisista analyysimenetelmistä
- datapalvelujen ekosysteemin luontia, kuten datan jakelupisteet, kauppapaikat ja julkaisukirjastot sekä datasovellusten edistäjärooli
- kokeilulaboratorioita, joissa voidaan testata teknologioita ilman omaa mittavaa panostusta

Keskeisten kansallisten toimijoiden tunnistaminen ja organisoimismallin luominen on tarpeen myös infrastruktuurin kehityksessä. Infrastruktuurikoordinaation keskeinen tehtävä on tuoda tehokkuutta, skaalaetuja ja yhteentoimivuutta tietojenkäsittelyn ja palveluiden kehitykseen. Esimerkkinä kansallisesta kumppaneille avoimesta ja kokeilut mahdollistavasta infrastruktuurista voidaan pitää esimerkiksi vuonna 2014 käyttöön otettua Digilen palvelupaja Forgeille.<sup>59</sup>

### Toimenpide-ehdotukset:

***Kehitetään kansallista big data infrastruktuuria verkostomaisella yhteistyöllä, jossa huomioidaan myös tutkimuksen ja opetuksen tarpeet.***

***Vahvistetaan huippuluokan kansainvälisiä tietoliikenne- ja viestintäyhteyksiä.***

***Tuetaan tietojenkäsittelyn tehostamista ja palveluiden kaupallistamista edistävien kansallisten infrastruktuurien kehittämistä julkisten ja yksityisten toimijoiden yhteistyöllä kokeilu- ja testausympäristöissä.***

## 6.5 Tekniset käytännöt ja standardit

Big datan teknologisia kehitystarpeina on tunnistettu erityisesti datan varastoinnin tekniset standardit sekä datan välittämisen ja integroinnin tavat. Teknologiakohtaisia ekosysteemejä on syntymässä, joten näiden ymmärtäminen ja sovittaminen omaan kehitysympäristöömme on olennaista. Datan käsittelyn ja hallinnoinnin harmonisointi edistää datan yhteen toimivuutta ja siten käyttöä. Dataan liittyviä käytäntöjä (esimerkiksi IN-

<sup>59</sup> Palvelupaja FORGE. <https://forageservicelab.fi/fi>

SPIRE-direktiivin<sup>60</sup> kautta) on jo olemassa, joita tulee soveltaa myös big datan käytön tehostamiseen.

Suomen ja suomalaisten yritysten tulee pysyä mukana ja edistää standardisointikehitystä osallistumalla mahdollisuuksiensa mukaan erilaisten ja eri alojen standardointielinten toimintaan. Lisäksi tietoa standardien ja teknologian kehityksestä tulee jakaa keskinäisessä verkostossa.

Avoimen lähdekoodin kehityksen seuranta ja siihen osallistumista tulee lisätä big datan teknologioiden ymmärtämiseksi (esimerkiksi 3WC, COSS yhteisöt). Tieto- ja yksityisyydensuoja sekä tietoturva ovat myös teknologisia kehityskysymyksiä, joiden ratkaisuun on panostettava.

#### **Toimenpide-ehdotukset:**

***Tunnistetaan keskeiset standardointiprosessit ja organisoidutaan toimimaan niissä seuraamalla alan kansainvälistä ja kansallista kehitystä.***

***Henkilötiedon hyödyntäminen tieto- ja yksityisyydensuoja säilyttäen on myös teknologinen kehityskysymys. Panostetaan läpinäkyvän ja turvallisen henkilötietoinfrastruktuurin teknologiakehitykseen.***

## **6.6 Datan käytettävyys ja avoimuus**

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

Hallintoon tulee saada big data -osaamista kehittäviä henkilöitä. Voidaan esimerkiksi rekrytoida osaajia erilaisiin projekteihin kumppanikoodarimallilla.<sup>61</sup>

Kunta- ja kaupunkitason kehitystyötä ja yhteishankkeita tietoaaineistojen jakamiseksi on edelleen edistettävä, sillä näillä on merkittävä osa hyödyllisestä datasta palvelukehitykselle.

#### **Toimenpide-ehdotukset:**

***Big datan hyödyntämiseen ratkaisuja etsiviä ja kehittäviä henkilöitä rekrytoidaan hallintoon.***

***Yhteentoimivuutta ja datan jakamista kehitetään yhteishankkeilla erityisesti kunta- ja kaupunkiympäristössä.***

<sup>60</sup> Euroopan parlamentin ja neuvoston direktiivi 2007/2/EY Euroopan yhteisön paikkatietoinfrastruktuurin (INSPIRE) perustamisesta. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:fi:PDF> . Kts. myös Paikkatietoikkuna: tietoa Inspire-direktiivistä. <http://www.paikkatietoikkuna.fi/web/fi/tietoa-inspiresta>

<sup>61</sup> Helsinki Region Infoshare: Kumppanikoodarit vievät open data -ilosanomaa virkamiehille. <http://www.hri.fi/fi/ajankohtaista/kumppanikoodarit-vievat-open-data-ilosanomaa-virkamiehille/>

***Huomioidaan kansallisen dataportaalin kehittämisessä tietoaineistojen jakelu- ja viestintäkanavana myös big datan mahdollisuudet.***

***Kannustetaan virastoja tunnistamaan data-aineistoja, joista big data kehityksen kannalta saataisiin merkittäviä hyötyjä. Tuetaan tässä virastojen pilottiprojekteja Avoimen tiedon ohjelman ja virastojen talouskehysten puitteissa.***

## 6.7 Sääntely

Suurena kehityshaasteena on tasapainottaa big datan hyödyntämiseen ja tietosuojaan liittyvät toimet sääntelyssä. Henkilötiedot ja yksityisyyden suoja eivät saa missään tilanteissa vaarantua.

Sääntelyn kehittämisen ajatuksena tulee kuitenkin olla se, että markkinoille tulon esteitä voidaan purkaa ja samalla luoda kannusteita datan saatavuuteen huomioiden henkilöiden oikeudet dataan. Yleisesti säädösvalmistelussa tulee huomioida datan edistykseen liittyvien edellytysten parantaminen.

Sääntelyä tukevat yhteiset käytäntösäännöt ja periaatteet. On luotava ns. "big data -etiketti" siitä, miten dataa käsitellään ja hyödynnetään kunnioittaen niin kansalaisten kuin yritystenkin oikeuksia. Hyviä datan hallinnon ja käsittelyn tapoja tulee luoda yhteistyössä julkisen ja yksityisen sektorin sekä yhteisöjen ja järjestöjen kanssa. Selkeät ja luotettavat käytäntösäännöt sekä korkea tietosuojatoimiva myös Suomen vahvuutena kansainvälisessä kilpailussa datavarannoista.

Erilaisten datan käytettävyyttä ja saatavuutta esimerkiksi tutkimuspuolella lisäävien avointen lisenssi- ja sopimusmallien kehitystä ja käyttöä tulee tukea.

### **Toimenpide-ehdotukset:**

***Otetaan huomioon kaikessa säädösvalmistelussa ja implementoinnissa vaikutukset big datan hyödyntämisen edellytyksiin.***

***Vaikutetaan siihen, että EU:n tietosuojasetus ja sen soveltaminen Suomessa mahdollistavat big data -kehityksen yksityisyydensuojaa vaarantamatta.***

***Kehitetään toimijoiden yhteistyönä käytäntösääntöjä, ns. suomalainen "big data -etiketti" erityisesti henkilötietojen käyttöön.***

## 6.8 Yhteistyö ja datan vaihto

Datan hyödyntämisen avulla on haettavissa synergioita eri aloilta. Yksittäisellä toimijalla ei välttämättä ole resursseja koota yhteistyöverkostoja. Kansallisesti on siten toteutettava toimia, joilla erityisesti yritysten horisontaalista ja vertikaalista verkottumista voidaan edesauttaa siten, että dataa jaetaan ja tarjotaan sopivilla alustoilla tai rajapinnoilla analyysia varten, sovelluskäyttöön tai yhdisteltäväksi uudella tavalla.

Haasteena nähdään myös se, ettei yrityksillä ole välttämättä kannusteita tai liiketoiminnan intressien kannalta järkevää tapaa jakaa dataansa. Riskinä on se, että käyttämättömästä datasta ei saada arvoa ja kilpailijat saavat tästä etua. Yhteistyön kehittäminen vaatii uutta ymmärrystä jakamistalouteen ja siten hyödyn- ja tulonjakomallien kehittämistä. Datan markkinapaikkoja tulee myös kehittää.

EU-tasolla panostetaan yhteistyömuotoihin big datan hyödyntämisessä. EU:ssa on perustettu "Big Data Public Private Forum"<sup>62</sup>, jossa luodaan yhteistyön malleja, fasilitoidaan datan siirtoa ja opitaan samalla miten liiketoiminta ja johtaminen muuttuu datalähtöiseksi. Foorumilla tehdään myös kokeiluja big data -aineistoilla. Niiden avulla kehitetään osaamista, tutkimuksen hyödyntämistä ja työkalujen siirtoa. Suomen tulee osallistua unionin laajuiseen big data -yhteistyöhön.

Osaavien ihmisten välisten yhteyksien rakentaminen katsotaan tässä kehitysvaiheessa olennaiseksi, samoin kuin hyvien käytäntöjen jakaminen eri toimijoiden välillä. Verkosto toiminnan tavoitteiden on oltava kuitenkin selkeitä: toiminnan on tuettava yritysten ja ihmisten suorituskykyä tuloksia seurattavissa olevalla tavalla.

Verkottuneet käyttäjät ja joukkoistaminen ovat myös big data -kehityksen keskiössä, sillä ne edesauttavat useilla sektoreilla tiedon hyödyntämisen nopeutumista ja tiedon rikastumista. Esimerkiksi liikenteessä reaaliaikaisesti verkkoon kytkeytyneiltä käyttäjiltä kerättyä tietoa voitaisiin yhdistää liikenteen hallinnan ratkaisuihin ja palveluihin.

#### **Toimenpide-ehdotukset:**

***Selvitetään tarvetta big data -hubille, erityisesti isojen ja pk-yritysten yhteisille datahankkeille ja datan vaihdolle liiketoiminnassa.***

***Edistetään erityisesti pk- ja start-up-yritysten big data -resursseja (välineitä ja pääsyä aineistoihin) ja osaamista. Aktivoidaan näitä yrityksiä myös mukaan kansallisiin hankkeisiin ja yritysten aloitteisiin.***

***Luodaan kansallinen data-analytiikan toimijoiden verkosto tutkimuslaitos-, yliopisto ja oppilaitosyhteistyössä, jolla edistetään myös teknologiasiirtoa yrityksiin.***

***Aktiivinen osallituminen ja kansallisten näkemysten jakaminen EU:n Big/NESSI –datafoorumeilla.***

***Kehitetään erityisesti joukkoistamiseen perustuvia big data -hyödyntämismalleja, joista parhaimmillaan syntyy uudentyyppisiä laajoja ekosysteemejä, yhteiskunnan toimivuutta ja käyttäjiä palvelevia ratkaisuja.***

## **6.9 Kokeilut ja T&K -rahoituksen suuntaaminen**

Kokeilevaa toimintaa big datan hyödyntämiseksi ja analyysimenetelmien soveltamiseksi käytännön tilanteisiin ja yhteiskunnallisiin haasteisiin sekä uuteen liiketoimintaan on

<sup>62</sup> Big Data Public Private Forum (BIG). <http://www.big-project.eu/>

edistettävä. Big datan kokeilut tulisi saada siten vahvemmin mukaan erilaisten ohjelmien ja kehityshankkeiden osaksi.

Tekesin ja Akatemian eri ohjelmien kautta saatavaa rahoitusta tulee kohdentaa enemmän big data -kehitykseen. Oleellista on saada myös yritykset mahdollisuuksiensa mukaan lisäämään kehitysinvestointeja big dataan ja analytiikkaan.

Tekes on käynnistänyt teollista internetiä koskevan ohjelmavalmistelu. Saman valmistelun yhteydessä tarkastellaan tarpeet myös 5G-mobiiliverkkojen alueella, joka on tärkeä tekijä teollisen internetin toteuttamisessa. Valmistelun tuloksena voi olla yksi tai useampi ohjelma tai jokin muu kokonaisuus. Valmistelun tuloksena mahdollisesti käynnistyvä(t) ohjelma(t) tai muut toimenpiteet voisivat käynnistyä syksyllä 2014 (Q4/2014).<sup>63</sup>

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

#### **Toimenpide-ehdotukset:**

***Kansallista tukea ja verkostoa EU-tukiohjelmien hyödyntämiseen vahvistetaan erityisesti big data -painotuksella.***

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

***Käynnistetään kokeilevia kehityshankkeita kaupallistamisen ja tulonjakomallien kehittämiseksi.***

***Pääomaehtoisen rahoituksen (yksityiset ja julkiset pääomasijoittajat) saatavuutta tulee edistää.***

#### **6.10 Omadata**

Omadatan edellytyksiä on kehitettävä, jotta yksilöillä olisi parempi kontrolli itseään koskevan tiedon hallitsemiseen ja mahdollisuus antaa tietoja palvelukehitykseen. Viranomaistoiminnassa henkilötietojen käsittelystä säädetään erikseen, eikä tietojen käyttö ja luovutus aina perustu henkilön omaan suostumukseen. Mydata-kehityksen yhteydessä tulee pohtia myös viranomaisten henkilörekistereissä olevan tiedon käsittelyperiaatteista.

***Omadata-selvityksen malleja testataan valituilla alueilla kehityksen vauhdittamiseksi ja uusien tietomallien yleistymiseksi eri alojen yritysten avulla.***

***Pitkällä aikavälillä luodaan omadatalle alustaa, johon yksilö voi kerätä ja halumallaan tavalla jakaa sekä hyödyntää omia tietojaan.***

<sup>63</sup> TEKES: Teollinen internet ja 5G. <http://www.tekes.fi/teollinen-internet-ja-5G>.

**Arvioidaan erikseen omadata- mallien soveltuvuutta viranomaisten toimintaan kansallisen palveluarkkitehtuurityön osana.**

**Yhteistyö ja verkostoituminen esimerkiksi Iso-Britannian MiData-aloitteen ja muiden maiden toimijoiden kanssa omadata-kehityksen vahvistamiseksi.**

## 7. Vaikuttavuus

Tässä osiossa arvioidaan big datan potentiaalia kansantaloudessa ja tuodaan esille näkökulmia siitä, miten raportissa esitettyjen toimenpiteiden vaikuttavuutta voidaan arvioida. Kuvaan 11 on tiivistetty joitakin big datasta eri kannalta saatavia hyötyjä.

**Kuva 11. Big datasta saatavia hyötyjä**

HALLINTO	YRITYKSET	KANSALAISET	TUTKIMUS
			
Prosessien tehokkuus Paremmat palvelut Läpinäkyvyys Kustannussäästöt ja resurssien kohdentuminen Päätöksenteon laatu	Innovaatiot Markkinaosuuksien tai voittojen kasvattaminen Prosessien tehokkuus (optimointi) Omaisuuden käyttö Tuottavuus Hävikin vähentäminen Asiakaskokemuksen parantaminen	Palvelujen laatu ja valintamahdollisuudet Uudet palvelut Parempi kontrolli itseä koskeviin tietoihin Vaikuttamismahdollisuudet Yhteiskunnallinen osallistuminen Kuluttajan oikeudet	Uusien aineistojen mahdollistamat laajemmat tutkimushankkeet Tiedon yhdistelystä uutta tutkimusta Menetelmistä tutkimuksen laatua ja tehokkuutta Tutkimuksen vaikuttavuus

© Johannes Jansson, Wikimedia Commons, Clipart

**Kansantalous:** Suurten tietoaaineistojen hyödyntämisen vaikuttavuuden ja taloudellisten tulosten laajempi arviointi on vielä varsin alussa. Akateemisia tutkimuksia vaikutuksista ja esimerkiksi suurten tietoaaineistojen hyödyntämisen laajemmasta kansantaloudellisesta merkityksestä on konsulttiyhtiöiden selvityksiä lukuun ottamatta vasta käynnisteillä<sup>64</sup>.

DemosEuropa, Varsovan taloustieteellinen instituutti ja Microsoft julkaisivat hiljattain "Big and Open data in Europe" -tutkimuksen, jonka mukaan koko EU:n alueella avoimen da-

<sup>64</sup> Llewelyn & Leiponen (2014), Commercializing big data: Emerging insights and a Research Agenda. Imperial College Lontoossa on aloittanut aiheeseen liittyvän tutkimushankkeen, jonka yhtenä vetäjänä toimii suomalainen Aija Leiponen. <http://www.imperial.ac.uk/business-school/research/innovation-and-entrepreneurship/ie-research/research-initiatives-and-themes/releasing-the-value-of-big-data/>



tan ja big datan laaja hyödyntäminen voi kasvattaa bruttokansantuotetta 1,9 % enemmän verrattuna tilanteeseen, jossa tietoaineistoja ei hyödynnetä. Vuoteen 2020 mennessä EU-alue voi tutkimuksen mukaan saavuttaa 200 miljardin euron hyödyt avoimen ja big datan hyödyntämisellä.<sup>65</sup>

Suomen osalta tutkimus arvioi big datan hyödyntämisen tuovan noin 2,1 % prosentin kasvun bruttokansantuotteeseen 2020 mennessä. Suomi luetaan potentiaaaliltaan parhaaseen pohjoiseen alueeseen. Bruttokansantuotteessa henkeä kohden tutkimuksessa laskeaan 560 euron lisäys.

Uuden teknologian ja ICT:n hyödyntämisen mahdollisuuksia mittaavassa WEF:in Networked Readiness Indexissä<sup>66</sup> Suomi on kärjessä. Indeksissä tieto- ja viestintäteknologian taloudelliset vaikutukset on eritelty yhdeksi alueeksi. Indeksissä mitataan muun muassa tieto- ja viestintäteknologian vaikutusta uusien palvelujen ja tuotteiden syntyyn. Suomella on arvioitu olevan parhaat mahdollisuudet tällä alueella, samoin kuin uusien organisaatiomallien kehityksessä.

**Hallinto:** Konsulttiyhtiö McKinseyn vuonna 2011 julkaisemassa tutkimuksessa "Big data: The next frontier for innovation, competition and productivity"<sup>67</sup> arvioidaan big datan hyödyntämisen potentiaalinen vuosittainen arvo Yhdysvaltojen terveydenhuollolle 300 miljardiksi dollariksi ja EU-alueen julkiselle sektorille 250 miljardiksi euroksi. Huomionarvoista on myös tutkimuksen lupaus uusista työpaikosta: pelkästään Yhdysvalloissa töitä on luvassa 140 000 – 190 000 data-analyttikolle.

**Yritykset:** Edellä mainitussa "Big & Open Data in Europe"- tutkimuksessa Suomen kannalta merkittävimmäksi ja potentiaalisimmaksi sektoriksi arvioitiin teollisuus, jonka osuus big datan hyödyntämiseen perustuvasta koko bruttokansantuotteen kasvusta arvioitiin noin puolen prosentin kokoiseksi. Ylipäänsä tutkimuksessa teollisuus nähtiin potentiaalisimmaksi alueeksi koko EU:n alueella. Suomen osalta muita tutkimuksia ei ole saatavilla.

Yleisesti ottaen datalähtöisesti päätöksiä tekevien yritysten on tutkittu olevan keskimäärin noin 5-6- prosenttia tuottavampia kuin perinteisillä menetelmillä päätöksiä tekevät yritykset.<sup>68</sup>

Tarkkaa analyysiä eri sektoreiden ja toimialojen big data -potentiaalista ei ole Suomessa tehty. Maailmalla tutkimuksia aiheesta on niin ikään varsin niukasti, muutamia on kuitenkin tehty. Ison Britannian (ja myöhemmin toistettuna Irlannin) osalta aihetta on tutkinut vuonna 2011 businessanalytiikkayhtiö SASin toimesta talouden ja liiketoiminnan tutkimuskeskus CEBR<sup>69</sup>.

<sup>65</sup> Buchholtz, Bukowski & Śniegocki (2014), Big & open data in Europe: A growth engine or a missed opportunity? [http://www.bigopendata.eu/wp-content/uploads/2014/01/bod\\_europe\\_2020\\_full\\_report\\_singlepage.pdf](http://www.bigopendata.eu/wp-content/uploads/2014/01/bod_europe_2020_full_report_singlepage.pdf)

<sup>66</sup> Ks esim. WEF, The Global Information Technology Report 2014: Rewards and Risks of Big Data.

[http://www3.weforum.org/docs/WEF\\_GlobalInformationTechnology\\_Report\\_2014.pdf](http://www3.weforum.org/docs/WEF_GlobalInformationTechnology_Report_2014.pdf)

<sup>67</sup> McKinsey Global Institute (2011), Big data: The next frontier for innovation, competition, and productivity.

[http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI\\_big\\_data\\_full\\_report.ashx](http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx)

<sup>68</sup> Brynjolfsson, Hitt & Kim (2011), Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?

<http://dx.doi.org/10.2139/ssrn.1819486>

<sup>69</sup> CEBR (2012): Data equity: Unlocking the value of big data. <http://www.sas.com/offices/europe/uk/downloads/data-equity-cebr.pdf>

Taulukossa 2 on esitelty CEBRin arviota eri alojen PK-yritysten tuottavuuden kasvusta vuoteen 2017 mennessä, jos big data -ratkaisuja otetaan käyttöön.

**Taulukko 2. Pk-sektorin yritysten tuottavuuden kasvu big datan hyödyntämisellä**

<b>Teollisuus</b>	<b>4,7 %</b>
<b>Vakuutus</b>	<b>4,1 %</b>
<b>Vähittäiskauppa</b>	<b>3,1 %</b>
<b>Energia</b>	<b>2,5 %</b>
<b>Investointipankit</b>	<b>2,4 %</b>
<b>Kuljetus- ja logistiikka</b>	<b>2,4 %</b>
<b>Telekommunikaatio</b>	<b>2,2 %</b>
<b>Hallinto</b>	<b>2,1 %</b>
<b>Muut alat</b>	<b>1,7 %</b>
<b>Pankit</b>	<b>1,6 %</b>
<b>Terveysala</b>	<b>1,5 %</b>
<b>”Professional services”</b>	<b>1,2 %</b>

Lähde: CEBR

**Tutkimus:** Tutkimuspuolella big data -kehitys mahdollistaa muun muassa uudenlaisten menetelmien käytön ja uudenlaisen tutkimuksen tekemisen. Teknologia- ja big data -kehityksen vaikuttavuutta tutkimuksen kannalta kuvaa hyvin esimerkki genomisen sekvensoinnin hintakehityksestä. Yhden genomisen sekvensoinnin hinta on tippunut vuoden 2001 noin 95 miljoonasta dollarista vuoden 2014 tammikuuhun mennessä noin 4000 dollariin<sup>70</sup>. Tutkimuksesta big data -menetelmien avulla saatavat hyödyt ovatkin erityisen suuria.

**Toimenpiteiden** käynnistämisen ja tarkemman toimenpidesuunnitelman laatimisen yhteydessä tulee paikata big dataan ja tiedon hyödyntämiseen liittyvää vaikutustutkimusten puutetta ja samalla määritellä tarkempaa arviota tarvittavista investoinneista. Ohessa on esitetty näkemyksiä siitä, miten vaikutuksia tulisi mitata:

<sup>70</sup> National Human Genome Research Institute: DNA sequencing costs (2014). <http://www.genome.gov/sequencingcosts/> kts. myös esim. O’Driscoll et al (2013): “Big data”, Hadoop and cloud computing in genomics. Journal of Biomedical Informatics 46:5, 774-781. <http://www.sciencedirect.com/science/article/pii/S1532046413001007>

- Esitetään arvioinnin pohjalta, miten ja kuinka merkittävästi edellä mainittujen toimien toteuttaminen vaikuttaisi yhteiskuntaan tai talouteen, esimerkiksi kestävyysvajeeseen.
- Etsitään kansantalouden tasolla esimerkkejä demonstroimaan big datan käytön vaikutuksia kasvuun ja työllisyyteen. Tämä vaatii myös datan relevanttiuden arviointia, kuten: 1) datan määrällisen kysynnän pohtimista ja 2) eri sektoreiden kansantaloudellista merkitystä big datan hyödyntämisen kannalta.
- Arvioidaan suomalaisen, eri alojen public private partnership -ekosysteemin edellytykset ja potentiaali. Ensin on ymmärrettävä vielä tarkemmin toimijoiden ja toiminnan laaja kirjo. Tässä raportissa ja sen lausuntokierroksella osaltaan tehtyä kartoitusta keskeisistä toimijoista ja hankkeista jatketaan.

#### **Toimenpide-ehdotukset:**

***Tehdään taloudellista ja yhteiskunnallista analyysia big datan toimijakentästä, potentiaalista Suomessa sekä strategisten toimenpiteiden vaikuttavuudesta.***

***Tehdään vaikuttavuusanalyysiä yhdessä Avoimen tiedon ohjelman avoimen datan vaikutusarvioinnin kanssa.***

## **8. Toimeenpano**

Luodaan toteuttamisen tiekartta luodaan yhteistyössä tiekartta toimien vastuista, etene- misjärjestyksestä, aikatauluista, tarvittavista resursseista ja mahdollisista riskeistä. Vahvistetaan strategian toimeenpano valtioneuvoston periaatepäätöksellä. Toimet, jotka voidaan käynnistää heti, resursoidaan ja organisoidaan välittömästi.

#### **Toimenpide-ehdotukset:**

***Ehdottujen kansallisten kehityslinjausten toimeenpanosta luodaan aikataulutettu, vastuutettu ja toimenpiteiden osalta priorisoitu suunnitelma syksyllä 2014. Ehdotettujen toimenpiteiden taloudelliset vaikutukset mukaan lukien valtion budjettitalouden määrärahatarpeet arvioidaan osana toimeenpanosuunnitelmaa.***

# Liitteet

## Liite 1.

### Big datan käyttö- työryhmän jäsenet\*

Taru Rastas, puheenjohtaja	Liikenne- ja viestintäministeriö
Emil Asp, sihteeri	Liikenne- ja viestintäministeriö
Ismo Kosonen	Liikenne- ja viestintäministeriö
Sami Niinimäki	Opetus- ja kulttuuriministeriö
Antti Eskola	Työ- ja elinkeinoministeriö
Anne Kauhanen-Simanainen	Valtiovarainministeriö
Marko Heikkinen	Tekes
Juha Latikka	Suomen Akatemia
Johanna Bragge	Aalto yliopisto
Jyrki Nummenmaa	Tampereen yliopisto
Tarja Riihisaari	Ilmatieteen laitos
Minna Ruckenstein	Kuluttajatutkimuskeskus
Matti Vakkuri	Tieto Oyj
Pekka Lehti	Valuemotive Oy
Veikko Hara	Rovio Entertainment Ltd
Ulla Kruhse-Lehtonen	Sanoma Oyj
Jani Pirkola	CyberLightning Oy
Immo Salo	Ivorio Oy
Markku Rauhamaa	Nokia Solutions and Networks Oy

\* Lisäksi raportin laadinnassa ja kirjoitustyössä ovat työryhmän ulkopuolelta auttaneet Markku Alanko, Ivorio, Marko Forsblom, LVM, Tuomas Nieminen, LVM, Jan Juslen, Liikennevirasto, Tommi Arola, Trafi, Alekski Kallio, CSC ja Aija Leiponen, Imperial College London

## Liite 2. Asettamiskirje



LIIKENNE- JA  
VIESTINTÄMINISTERIÖ

11.12.2013

**Asia** Big datan käyttö – työryhmän asettaminen ("Big Suomi")

### Asettaminen

Liikenne- ja viestintäministeriö on asettanut työryhmän valmistelemaan big datan hyödyntämistä ja kehittymistä edistäviä toimia.

### Toimikausi

1.1.2014 – 30.6.2014.

### Tausta

Big dataa, jolla tässä yhteydessä tarkoitetaan kykyä hyödyntää suuria ja monimuotoisia digitaalisia tietoaaineistoja, pidetään yhtenä taloudellisen kasvun avainteknologiana (OECD, UK, USA). Arviot big datan markkinoiden arvosta vaihtelevat 15-50 miljardin euron välillä.

Aihe on noussut KIDE ohjelmassa avoimen datan teeman rinnalle. Big data asiaa on esitelty järjestetyissä tilaisuuksissa keväällä ja syksyllä 2013. Kesällä 2013 teetettiin selvitys big data - markkinasta ja tarpeista Suomessa. Big data on ICT2015-toimenpidealue.

Esitetyn työryhmän tarkoitus on luoda näkemys kansallisista edistämistoimista, joilla parantaa edellytyksiä eri lähteistä saatavan datan tehokkaaseen ja innovatiiviseen käyttöön. Seurauksena tavoitellaan sitä, että tietoaaineistoja hyödynnetään paremmin yhteiskunnan eri sektoreilla.

### Tehtävät

Työryhmän tehtävänä on luoda kokonaiskuvaa big datan tilanteesta Suomessa sekä tehdä ehdotuksia tarvittavista kehittämistoimista:

Data-aineistojen käytettäväksi saaminen

- kartoittaa erilaisten lähdetietojen käytön mahdollisuuksia
- tunnistaa käytön esteitä ja käytön helpottamiseen tähtäävien teknisten, sääntelyn tai hallinnollisten uudistusten tarpeita

Data-aineistojen hyödyntäminen eri sektoreilla

- määrittää keskeisiä ja potentiaalisia soveltamisalueita ja kehittämistoimintoja
- vaikuttaa käyttöä edistävien toimintatapojen kehittämiseen ja pilotointiin

Aiheeseen liittyvä osaaminen ja tutkimus

- tehdä ehdotuksia osaamisen kehittämiseksi
- tunnistaa vaikutusalueita ja tutkimustarpeita

Lopputuotoksena työryhmä valmistelee luonnoksen big data strategiaksi kansallisina kehittämislinjauksina.

Työryhmän tulee saada työnsä valmiiksi 30. kesäkuuta 2014 mennessä.

### **Organisointi**

Työryhmän kokoonpano:

Puheenjohtaja

Taru Rastas, liikenne ja viestintäministeriö

Sihteeri

Emil Asp, liikenne- ja viestintäministeriö

Jäsenet

Ismo Kosonen, LVM

Juha Haataja, OKM

Antti Eskola, TEM

Anne Kauhanen-Simanainen, VM

Marko Heikkinen, Tekes

Juha Latikka, Suomen Akatemia

Johanna Bragge, Aalto yliopisto

Jyrki Nummenmaa, Tampereen yliopisto

Tarja Riihisaari, Ilmatieteen laitos

Jussi Melkas, Tilastokeskus

Minna Ruckenstein, Kuluttajatutkimuskeskus

Matti Vakkuri Tieto Oy/ICT2015

Pekka Lehti, Valuemotive Oy

Veikko Hara, Rovio Entertainment Ltd

Ulla Kruhse-Lehtonen, Sanoma Oy

Jani Pirkola, Cyber Lightning Oy

Immo Salo, Ivorio Oy

Markku Rauhamaa, NSN

Koska työryhmän jäsenet ja sihteeri on valittu erityisasiantuntemuksensa perusteella, naisten ja miesten välisestä tasa-arvosta annetun lain säännöksiä työryhmän kokoonpanosta ei ole voitu noudattaa.

Työ tehdään virkatyönä. Työryhmä voi käyttää tehtävässään tarvittaessa konsulttityötä tai teettää selvityksiä. Työryhmä toimii yhteistyössä asianosaisten toimijoiden kanssa.

### Liite 3

#### Sanasto ja suomennosehdotuksia

Ehdotukset Kotimaisten kielten tutkimuslaitokselta

<b>Big data</b>	massadata; iso data (TSK: Tietotekniikan termi- talkoot <sup>1</sup> )
<b>Mydata:</b>	omadata
<b>Data-driven</b>	tietolähtöinen, datalähtöinen
<b>Quantified self</b>	itsensä mittaaminen
<b>Massive online open course</b>	laajalle joukolle järjestettävä verkkokurssi, kaikille avoin verkkokurssi
<b>Data scientist</b>	datatieteilijä
<b>Machine-to-machine</b>	koneiden välinen viestintä
<b>Internet of things</b>	esineiden internet (TSK)
<b>De-identification</b>	tunnistustietojen poistaminen
<b>Data retention &amp; non-retention</b>	datan säilyttäminen & tuhoaminen
<b>Industrial internet</b>	teollinen internet
<b>Cleantech</b>	puhdas teknologia

---

<sup>1</sup> <http://www.tsk.fi/tsk/termitalkoot/fi/etusivu-263.html>

#### **Liite 4:**

#### **Suomalaisten korkeakoulujen nykyistä ja käynnistymässä olevaa big data -koulutusta ja koulutusohjelmia**

Raportti lähetettiin lausunnoille myös kaikkiin Suomen korkeakouluihin. Tiedot big data -koulutuksesta on kerätty lausuntojen perusteella. Lista ei ole täysin kattava.

#### **Helsingin yliopisto:**

- Tietojenkäsittelytieteen laitos<sup>1</sup> käynnistää 2014-2015 uuden erikoistumisprofiilin maisterinopintoihinsa, jossa voi erikoistua datatieteeseen ja big dataan. Profiili yhdistää data-analyysin ja nykyaikaisen Internet-laskentaympäristön taidot.
- Tietojenkäsittelytieteen koulutuksessa opetusta big dataan liittyen mm. matemaattisesta mallinnuksesta ja hajautetusta laskennasta.
- Valtiotieteellisen tiedekunnan tohtorikoulutuksessa huomioitu big data- teema.
- Nykykielten laitoksella laajojen tekstiainestojen tutkimuskäytön opetusta korpuslingvistiikassa ja kieliteknologiassa.
- Humanistiseen tiedekuntaan perustettu digitaalisten aineistojen professuuri.<sup>2</sup>

#### **Aalto-yliopisto:**

- Kauppakorkeakoulussa maisteriohjelma Information and Service Management<sup>3</sup>, suuntautumisvaihtoehtona Business Analytics. Suoraan big data -koulutusta.
- 2014 alkaen kaikkiin maisteriohjelmiin mahdollista sisällyttää sivuaine Analytics and Data Science<sup>4</sup>.
- Syksyllä 2015 alkavat Aallon uudistetut maisteriohjelmat. ICT-alan ohjelmassa Tieto-, tietoliikenne- ja informaatiotekniikka on tietotekniikan hakukohteessa erikoistumisvaihtoehtona Big Data and Large Scale Computing
- Maisteriohjelma Foundations of Advanced Computing<sup>5</sup>.
- Maisteriohjelma Master's Programme in Machine Learning and Data Mining.<sup>6</sup>

#### **Lappeenrannan teknillinen yliopisto:**

- Kauppätieteen kandidaatin tutkinnossa pakollisia data-analyysikursseja. Näissä tullaan soveltamaan myös data mining -työkaluja (SAS Enterprise Miner)
- Kauppätieteen maisteriohjelma Master's Degree in Strategic Finance and Business Analytics<sup>7</sup>, tässä ja muissa kauppätieteen maisteriohjelmissa isojen aineistojen analyysiin liittyviä soveltuvia harjoitustöitä ja projekteja yritysten kanssa.
- Tuotantotalouden ohjelma on uusimassa innovaatiojohtamisen pääainettaan. Mahdollisuus kehittää ohjelmasta big data -osakokonaisuus. Lisäksi tuotantotalouden ja tietotekniikan opintojen yhdistämisellä voidaan tuottaa valmiuksia big dataan liittyvälle työlle.

#### **Tampereen teknillinen yliopisto:**

- Data Engineering -pääaine tietotekniikan diplomi-insinöörin tutkinto-ohjelmassa.
- TTY:n tavoitteena on big data -tutkimuskeskuksen perustaminen vuoden 2016 aikana, panostuksia jo laitettu alueelle.

---

<sup>1</sup> <http://www.cs.helsinki.fi/opiskelu>

<sup>2</sup> <http://www.helsinki.fi/hum/ajankohtaista/2013/12/1216c.htm>

<sup>3</sup> [http://www.aalto.fi/fi/studies/education/programme/information\\_service\\_management\\_master/](http://www.aalto.fi/fi/studies/education/programme/information_service_management_master/)

<sup>4</sup> <http://users.ics.aalto.fi/gionis/AnalyticsDataScience.pdf>

<sup>5</sup> [http://www.aalto.fi/fi/studies/education/programme/foundations\\_of\\_advanced\\_computing\\_master/](http://www.aalto.fi/fi/studies/education/programme/foundations_of_advanced_computing_master/)

<sup>6</sup> [http://www.aalto.fi/fi/studies/education/programme/machine\\_learning\\_data\\_mining\\_master/](http://www.aalto.fi/fi/studies/education/programme/machine_learning_data_mining_master/)

<sup>7</sup> <https://uni.lut.fi/en/web/lut.fi-eng/master-s-degree-programme-in-strategic-finance-msf>



- TTY:n täydennyskoulutuskeskus Edutech järjestää "Big data ja data-analytiikka liiketoiminnan kehittämisessä"- täydennyskoulutusta.<sup>8</sup>

### Tampereen yliopisto:

- International Master's Programme in Computational Big Data Analytics, tilastotieteen ja tietojenkäsittelytieteen vetämänä, yhteistyössä eri alojen kanssa (mm. bioinformatiikka, biometria, biostatistiikka, yhteiskuntatieteet, datajournalismi). Alkaa syksyllä 2015.
- Perusteilla Tilastotieteen keskus, joka toimii eri alojen yhteistyöelimenä suurten tietoaineistojen hyödyntämisessä ja osallistuu tilastotieteellisten opetustavoitteiden laadintaan kaikissa yliopiston tutkinto-ohjelmissa.

### Arcada

- Maisterinohjelma International Business Management -A Nordic Focus<sup>9</sup>, jossa digitaalilla markkinoinnilla ja big data yhä vahvemmin esillä.
- Perusteilla myös big data -analyysiin keskittyvä informaatioanalyysin insinööriohjelma

### Itä-Suomen yliopisto:

- Tietojenkäsittelytieteen laitoksella mahdollista suorittaa pitkälti (big) dataan keskittyviä opintoja<sup>10</sup>. Mm. kehittynyt data-analytiikka johtamisessa -kurssikokonaisuus. (25op) Tietojenkäsittelytieteen koulutuskokonaisuuksissa hyödynnetään big data -tekniikoita, erityisesti Älykäs medialaskenta -kokonaisuudessa<sup>11</sup>.
- Kansainvälisen IMPIT-maisteriohjelman<sup>12</sup> yhtenä aihealueena big data.
- Big data -menetelmiä ja opetusta myös muilla aloilla: ympäristöinformatiikka (mm. Big GIS data-projektit), biolääketiede, bioinformatiikka, fysiikka, soveltava fysiikka ja matematiikka, tilastotiede.
- Pohjois-Savon yrityksille järjestetty marraskuussa 2013 ja helmikuussa 2014 data-analytiikan teemakoulutuksia<sup>13</sup>.

### Jyväskylän yliopisto<sup>14</sup>

- Tietotekniikan, matematiikan ja tilastotieteen laitosten yhteinen Data-analyysin monitieteellinen maisterikoulutus (DATA).<sup>15</sup>
- Informaatioteknologian tiedekunta kehittänyt data-analyysin koulutusta<sup>16</sup> muiden laitosten sekä Tel Avivin ja Yalen yliopistojen kanssa.
- Tilastotieteen maisteritutkinto, sovelletun matematiikan maisteritutkinto ja laskennallisten tieteiden maisteritutkinnoissa mahdollisuus profiloitua big data -analyysiin.<sup>17</sup>
- Keski-Suomen ICT-alan organisaatioille suunnatussa PROFIT-koulutuksessa käsitelty myös big dataa.<sup>18</sup>

<sup>8</sup> <http://www.tut.fi/fi/yrityksille/taydennyskoulutus/koulutustarjonta/big-data/index.htm>

<sup>9</sup> <http://www.arcada.fi/fi/node/1536>

<sup>10</sup> <http://www.uef.fi/fi/cs/datatiede>

<sup>11</sup> <http://www.uef.fi/fi/cs/alykas-medialaskenta>

<sup>12</sup> <http://cs.joensuu.fi/impit/>

<sup>13</sup> <http://www.uef.fi/fi/tietava/teemakoulutus>

<sup>14</sup> <https://www.jyu.fi/it/bigdata>

<sup>15</sup> <https://www.jyu.fi/it/laitokset/mit/opiskelu/maisteriopinnot>

<sup>16</sup> [https://www.jyu.fi/it/laitokset/mit/opiskelu/maisteriopinnot/MAISTERIKOULUTUS\\_TIEDOTE\\_22.4.13.pdf](https://www.jyu.fi/it/laitokset/mit/opiskelu/maisteriopinnot/MAISTERIKOULUTUS_TIEDOTE_22.4.13.pdf),

[https://www.jyu.fi/it/uutiset/tiedekunta/ITK\\_White\\_Paper\\_240614.pdf](https://www.jyu.fi/it/uutiset/tiedekunta/ITK_White_Paper_240614.pdf)

<sup>17</sup> [https://www.jyu.fi/it/uutiset/tiedekunta/dokumentit/KyberBD\\_1952014.pdf](https://www.jyu.fi/it/uutiset/tiedekunta/dokumentit/KyberBD_1952014.pdf)

## Haaga-Helia

- Business Information Technology-koulutusohjelmaan kuuluu big dataan liittyviä opintoja<sup>19</sup>.

---

<sup>18</sup> [https://www.jyu.fi/it/uutiset/tiedekunta/big\\_data\\_seminaari](https://www.jyu.fi/it/uutiset/tiedekunta/big_data_seminaari)

<sup>19</sup> <http://www.haaga-helia.fi/fi/opinto-opas/opintoiaxsokuvaukset/BUS8TF100> , <http://www.haaga-helia.fi/en/education/bachelor-degree-programmes/degree-programme-business-information-technology>