



Liikenne- ja
viestintäministeriö

Big data Suomessa

Keskustelualoite

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 päivämäärä 13.09.2013	
Julkaisun nimi Big data Suomessa	
Tekijät Markku Alanko, Ivorio Oy Immo Salo, Ivorio Oy	
Toimeksiantaja ja asettamispäivämäärä Liikenne- ja viestintäministeriö 30.05.2013	
Julkaisusarjan nimi ja numero Liikenne- ja viestintäministeriön julkaisuja 25/2013	ISSN (verkkójulkaisu) 1795-4045 ISBN (verkkójulkaisu) 978-952-243-358-9 URN http://urn.fi/URN:ISBN:978-952-243-358-9 HARE-numero
Asiasanat big data, tietovarannot, lähdedata, avoin data, tietojenkäsittely	
Yhteyshenkilö Taru Rastas p. 02953421617	
Tiivistelmä Selvityksen tarkoituksena oli kartoittaa eri sektoreiden näkemyksiä, tämän hetken tilannetta sekä tulevaisuuden odotuksia ja suunnitelmia liittyen meneillään olevaan big data ilmiöön. Big datalla tarkoitetaan yleisesti suurta määrää dataa, joka on muodoltaan vaihtelevaa. Tiedon keräämiseksi ja ymmärryksen lisäämiseksi erilliset kyselylomakkeet laadittiin ja lähetettiin julkishallinnon ja yksityisen sektorin toimijoille, keskeisimmille Suomessa toimiville palveluntarjoajille, konsulteille sekä koulutus- ja tutkimussektorin edustajille. Selvityksen rungon muodostavat kyselyn tulokset. Pohjatietona on käytetty asiantuntijoiden ja haastateltujen tietämystä ja näkemyksiä. Selvityksestä käy esille selkeästi kaksi seikkaa. Ensinnäkin big data on monelle organisaatiolle vielä uusi ja vaikeasti lähestyttävä asia. Toisaalta se kuitenkin tunnistettiin yleisesti strategiseksi muutosvoimaksi ja tulevaisuuden kilpailukyvyyn edellytykseksi. Tutkimus- ja koulutuspanostuksia ollaan lisäämässä ja budjetteja suuntaamassa hankkeille lähitulevaisuudessa. Ilmiönä big data koettiin lähes poikkeuksetta kiinnostavana ja merkittävänä. Merkittävimpinä haasteina big datan jalkautumisessa jokapäiväisen tietojenkäsittelyn ytimeen pidettiin kasvavaa osaajapulaa, ilmiöön liittyvän palveluja ohjelmistotarjonnan valtavaa kirjoa sekä lähdedatan räjähdysmäistä kasvua ja monimuotoistumista. Selvityksen johtopäätöksenä on, että big data ilmiöön on selvästi havahduttu ja meneillään olevan muutoksen merkittävyyden suuruudesta ollaan melko yksimielisiä. Se, mikä rooli big datalla ilmiönä on organisaation omalle toiminnalle, vaihteli sektoreittain. Pisimmälle meneviä täsmällisiä visioita oli palveluntarjoajilla. Koulutuspuolen alkavat hankkeet pyrkivät vastaamaan orastavaan osaajapulaan. Julkishallinnon rooli korostui erityisesti avoimen datan kautta dataliiketoiminnan mahdollistajana. Big datan ympärillä käytävää keskustelua on syytä jatkaa. Tiedon vallankumous avaa uusia mahdollisuuksia kaikille sektoreille. Suomen näkökulmasta tämä on ennen kaikkea mahdollisuus toimia kehityksen eturintamassa ja kehittää edelleen kilpailukykyämme tietojenkäsittelyn esimerkkimaana.	

Publikationsdatum 13.09.2013	
Publikation Big data i Finland	
Författare Markku Alanko, Ivorio Oy Immo Salo, Ivorio Oy	
Tillsatt av och datum Kommunikationsministeriet 30.05.2013	
Publikationsseriens namn och nummer Kommunikationsministeriets publikationer 25/2013	ISSN (webbpublikation) 1795-4045 ISBN (webbpublikation) 978-952-243-358-9 URN http://urn.fi/URN: ISBN: 978-952-243-358-9 HARE-nummer
Ämnesord big data, datalager, källdata, öppna data, databehandling	
Kontaktperson Taru Rastas tel. 02953421617	Rapportens språk Finska
<p>Sammandrag</p> <p>Syftet med utredningen var att belysa fenomenet big data ur olika sektors synvinkel samt att beskriva nuläget, framtida planer och förväntningar. Big data står för digitalt lagrade mycket stora datamängder i varierande form.</p> <p>För att samla in information om ämnet utarbetades separata frågeformulär för den offentliga förvaltningen och den privata sektorn, de viktigaste serviceleverantörerna i Finland, konsulter samt företrädare för utbildning och forskning. Stommen i utredningen utgörs av svaren på förfrågningen. Som underlag användes kunskap och åsikter framförda av experter och intervjuade serviceleverantörer.</p> <p>Utredningen påvisar tydligt två omständigheter. För det första är big data än så länge ett nytt och svårhanterligt begrepp för många organisationer. Å andra sidan erkänns big data allmänt som en strategisk förändringsfaktor och en förutsättning för framtida konkurrenskraft. Nu ökar satsningarna på forskning och utveckling och inom en snar framtid riktas anslag till big data-projekt.</p> <p>Big data bedömdes nästan utan undantag vara ett intressant och viktigt fenomen. De största utmaningarna för att integrera big data i de dagliga databehandlingsrutinerna ansågs vara den tilltagande bristen på experter, det brokiga urvalet tjänster och program samt den explosionsartade tillväxten och mångfalden av källdata.</p> <p>Slutsatsen av utredningen är att big data helt klart är ett fenomen som har noterats och att följderna av den pågående förändringen väntas bli betydande. Åsikterna om vilken roll big data som fenomen kommer att ha för enskilda organisationer varierade enligt sektor. De mest avancerade visionerna framfördes av serviceleverantörer. För att råda bot på den ökande bristen på experter startas nya utbildningsprojekt. Den offentliga förvaltningens roll är främst att skapa förutsättningar för den affärsverksamhet som öppna data erbjuder.</p> <p>Det är skäl att sporra diskussionen kring big data. Informationsrevolutionen öppnar nya möjligheter för alla sektorer. Ur finländskt perspektiv innebär detta framför allt en chans att inta en plats i tåten av utvecklingen och att vidareutveckla vår konkurrenskraft som en föregångare inom databehandling.</p>	



	Date 13.09.2013
Title of publication Big Data in Finland	
Author(s) Markku Alanko, Ivorio Oy Immo Salo, Ivorio Oy	
Commissioned by, date Ministry of Transport and Communications, 30.05.2013	
Publication series and number Publications of the Ministry of Transport and Communications 25/2013	ISSN (online) 1795-4045 ISBN (online) 978-952-243-358-9 URN http://urn.fi/URN:ISBN:978-952-243-358-9 Reference number
Keywords Big data, data reserves, source data, open data, data processing	
Contact person Taru Rastas tel. 02953421617	Language of the report Finnish
<p>Abstract</p> <p>The purpose of this report was to explore the current Big Data phenomenon with regard to how it is perceived and how it is expected to develop. Big Data refers to groups of data that are large and vary in form.</p> <p>In order to collect information about the phenomenon, questionnaires were drawn up and sent to customers struggling with increasing amounts of data both in the public and the private sector, to main service providers in Finland, to consultants, and to representatives in the research sector. Answers to the questionnaire form the main body of the report. The report is based on the knowledge and views of experts and interviewed service providers.</p> <p>The report clearly points out two issues. On the one hand, for many organisations Big Data is a new concept that is difficult to grasp. On the other hand, it was generally recognised as a strategic force of change and a prerequisite for future competitiveness. In the near future, investments in research and training will increase and new projects will be launched.</p> <p>Almost without exception Big Data as a phenomenon was considered interesting and significant. An increasing shortage of experts, enormous diversity of software services related to the phenomenon, and explosive growth and diversification of source data were considered the most significant challenges in bringing Big Data to the core of everyday data processing.</p> <p>The report concludes that the Big Data phenomenon has clearly been identified and there is unanimity over the extent and importance of the ongoing change. The role of Big Data to a single organisation varied by sectors. The most detailed visions were offered by the service providers.</p> <p>New training projects aim to respond to the growing lack of experts. The role of the public administration was emphasised as a provider of open data, which in turn would provide opportunities for data business.</p> <p>Debate about Big Data should be encouraged. Information revolution will open up new opportunities in all sectors. From Finland's point of view this is first and foremost an opportunity to be at the forefront of development, further improve our competitiveness and set an example in data processing.</p>	

Sisällysluettelo

1.	Johdanto	3
1.1	Kansainväliset big data –markkinat	4
1.2	Big data -selvityksen lähtökohdat	5
2.	Tavoitteet ja menetelmät	6
2.1	Internet-kyselyt	6
2.2	Haastattelut.....	6
2.3	Yksilöitävyys.....	6
3.	Big datan ja keskeisimpien teknologioiden esittely.....	7
3.1	Hadoop	7
3.2	Muistinvarainen analytiikka.....	7
3.3	Pilvipalvelut	7
4.	Jo tunnistetut keskeiset tiedontarpeet ja kehityskohteet.....	9
4.1	Tiedon jakaminen.....	9
4.2	Datalähteiden tunnistaminen ja hyödyntäminen	9
4.3	Oikeudelliset asiat	10
5.	Kyselyn roolien kuvaukset	11
5.1	Asiakkaat	11
5.2	Koulutus ja tutkimus.....	11
5.3	Julkishallinto.....	11
5.4	Konsultit	11
5.5	Palveluntarjoajat	12
6.	Selvityksen tulokset.....	13
6.1	Haastattelut.....	13
6.2	Verkkokysely	14
6.2.1	Yleiset kysymykset.....	15
6.2.2	Roolikohtaiset kysymykset: asiakkaat.....	18
6.2.3	Roolikohtaiset kysymykset: koulutus ja tutkimus	20
6.2.4	Roolikohtaiset kysymykset: julkishallinto.....	23
6.2.5	Roolikohtaiset kysymykset: konsultit	24
6.2.6	Roolikohtaiset kysymykset: palveluntarjoajat	26
7.	Selvityksen keskeiset havainnot	28
8.	Jatkotoimenpide-ehdotukset	29
8.1	Koulutus	29
8.2	Yritysten tukeminen.....	29
8.3	Julkishallinto kehityksen vauhdittajana.....	29
8.4	Yritysten innovaatiot ja yhteistyö	30
8.5	Palveluntarjoajien lokaali aktivoituminen	30
9.	Suomen data -foorumi	31
10.	Loppusanat.....	31

Esipuhe

Dataa syntyy useista lähteistä yhä kiihtyvällä tahdilla tallennettavaksi ja jaettavaksi teknisillä laitteilla hetkessä ympäri maailmaa. Analysoimalla data-aineistoja liiketoiminnan ja yhteiskunnan tarpeista lähtien on jo mahdollista ratkoa monimutkaisia haasteita. Yritykset näkevät digitaalisen tiedon hyödyntämisen merkittävänä nosteenä tuottavuudelle. Kyky käyttää tietoa toiminnan ajurina voi kuitenkin olla vajavaista, jos organisaation strategia ja osaaminen eivät taivu digitaaliseen ympäristöön.

Tämä selvitys big datasta on toteutettu Kide-ohjelman puitteissa. Ohjelman tavoite on löytää uutta kehitystoimintaa pilotoimalla ja digitaalisista palveluista kasvua. Selvityksen tarkoitus on kartoittaa big datan toimialaa ja kehittämistarpeita Suomessa sekä käynnistää keskustelua ja verkostomaista toimintaa tiedon hyödyntämisen vauhdittamiseksi yhteiskunnan eri sektoreilla.

Big data liittyy tiedon käytön näkökulmasta läheisesti avoimen datan edistämiseen. Tietoaineistojen vapaalla saatavuudella tavoitellaan innovatiivisten sovellusten lisäksi yhteiskunnallista tehokkuutta ja läpinäkyvyyttä. Tiedon käytön edistäminen on tarkoituksenmukaista, jos tietomassoista voidaan jalostaa hyvinvointia. Teknisesti digitaaliset ratkaisut ovat jo periaatteessa kaikkien ulottuvilla, joten kehityksen esteenä ovat lähinnä valmiudet toimia.

Raja datan ja tiedon välillä voidaan nähdä siinä kohtaa, jossa ihminen antaa merkityksen havaitulle asialle. Esimerkiksi puhelinten sensori- ja paikannusdatasta, tietokannoissa olevasta biopankki- ja rekisteridatasta tai verkkokauppojen hinta- ja ostotapahtumadatasta, sosiaalisen median sisältö- ja videotallennedatasta tai ilmakehän mittaus- ja sääasemien havaintodatasta tulee tietoa, kun sitä jäsenetään ja esitetään soveltuvasti. Toimintaan ohjaavien monimutkaisten ohjelmistojen, laskentamallien tai ratkaisujen takaa löytyy puolestaan laaja kirjo osaamisen yhdistämistä: koodivelhoa, datamatemaattikkaa, tilastonikkaria, kehittäjää, liiketoiminta-analytiikkaa, käyttäytymistieteilijää ja trendihaistajaa. Koneet välittävät keskenään dataa, mutta toimivaa älyä syntyy vasta, kun tähän kommunikointiin kytkeytyy ihminen.

Sen sijaan, että pelkona olisi internet-vallankumouksen seurauksena automatisoinnin hävittämät työpaikat, on paremminkin tiedon hyödyntämisen kyvyt ja näkemyksellisyys nousemassa rajoitteiksi kasvulle. Tarvitsemme siten lisää edellä mainittuja tiedon tekijöitä, jotta big datan hyödyt saadaan kaikkien yhteiskunnan sektorien käyttöön.

Myös luottamuksesta tulee datan saatavuuden myötä entistä tärkeämpää aineetonta pääomaa. Haasteena on luoda tiedon käytön pelisäännöt, jotka ovat tasapainossa odotettujen hyötyjen ja sovellettavan tietosuojan välillä. Kansainvälisten regiimien puitteissa käydään kilpaa datavarannoista, jossa edistyneet tiedon hallintatavat voidaan nähdä etuna.

Kuten tämä selvitys antaa suuntaa, kehittämispaineita on Suomessa erityisesti osaamisen ja toiminnan resursoinnin alueilla. Kokonaisuuden tasolla tarvitsemme panostuksia yhdistää internet big datan, mobiliteetin ja pilvipalveluiden potentiaaliin. Tämä tapahtuu niin koulutuksen ja riskirahoituksen suuntaamisella kuin yleisten toimintaedellytysten tasolla siten, että Suomi on hyvä paikka tietointensiiviselle yritteliäisyydelle. Toivomme, että tämä selvitys ja sen myötä rakentuva ideoiden jakaminen vievät asiaa askeleita eteenpäin.

Taru Rastas
Liikenne- ja viestintäministeriö

Johdanto

Kiinnostus big dataa kohtaan on tällä hetkellä valtava. Aiheesta puhutaan jossain muodossa niin yrityksissä, julkishallinnossa kuin aamupalapöydissä. Kattokäsitteenä se on niin laaja, että moni nykyajan ilmiö voidaan nähdä tiiviisti sen piiriin kuuluvana. Juuri laajuutensa vuoksi keskustelu big datasta on edelleen usein enimmäkseen teoreettista. Tämä selvitys pyrkii osaltaan tarjoamaan koosteen tähän kiehtovaan ilmiöön useista näkökulmista.

Suomen rooli teknologisten innovaatioiden kärkimaana on hiipunut viime vuosina. Big datan suhteen koko Eurooppa näyttää olevan vaarassa jäädä terävimmän kärjen ulkopuolelle. Vielä ei ole kuitenkaan myöhäistä reagoida. Vahva perusosaaminen tietojenkäsittelyssä, korkea koulutustasomme ja julkishallinnon aktiivinen rooli muodostavat vahvan perustan, jonka päälle innovaatioita ja uutta kilpailukykyä voidaan alkaa rakentamaan.

“Volume, velocity, variety”

Big data ilmiönä on kasvattanut tunnettuuttaan eksponentiaalisesti vuodesta 2011 lähtien. Itse käsite tuli tunnetuksi jo huomattavasti aikaisemmin ja esimerkiksi usein big dataa luonnehtimaan käytetyt kolme V-kirjainta (Volume, Velocity, Variety eli volyyymi, vauhti ja variaatio) esiintyivät Gartnerin raportissa jo vuonna 2001. Ilmiön tunnettuuden radikaali kasvu ajoittuu kuitenkin 2010-luvulle.

Big datalle ei ole yhtä tyhjentävää määritelmää. Se, mistä näkökulmasta aihetta kulloinkin lähestytään, riippuu lähestyjän intresseistä ja taustasta. Pitkään markkinoilla olleet yritykset, jotka myyvät datan säilyttämiseen, siirtämiseen, turvaamiseen ja analysointiin liittyviä palveluita, korostavat Big dataan liittämensä tuotteiden, ohjelmistojen ja palveluiden yhteydessä jatkuvuutta sekä integroitavuutta jo aiemmin tarjoamiinsa ratkaisuihin. Uudet startup-tyyppiset yritykset sekä monet konsultit puolestaan painottavat epäjatkuvuutta ja radikaalin muutoksen voimaa. Ne yritykset puolestaan, jotka eivät markkinointiviestinnässään käytä käsitettä Big data, koska se ei edes etäisesti sovi olemassa olevaan tarjoomaan, monesti vähättelevät koko käsitettä tai leimaavat sen ohi meneväksi hypeksi, johon ei vakavasti liiketoimintaansa suhtautuvan asiakkaan kannata kiinnittää huomiota. Tämä on ymmärrettävää oman liiketoiminnan ollessa uhattuna. Vastaavaa toimintamallia sovelsivat useat yritykset pilvipalveluiden alkutaipaleella, kun ei omaa tarjottavaa markkinoille vielä ollut.

Lähde: Big Data - tiedon vallankumous, Docendo 2013

Yksi käytetyimmistä hahmotelmista big data -käsitteen ymmärtämiseksi on edellä mainittu kolmen V-kirjaimen malli, jota on täydennetty myös muilla V-kirjaimella alkavilla ilmiötä kuvaavilla substantiiveilla.

Volyyymi

Datan määrä maailmassa ilmaistaan tällä hetkellä olevan muutaman zettatavun (10^{21} tavua) luokkaa ja kasvun ennustetaan olevan vuositasolla kymmeniä prosentteja. Tutkimuslaitos IDC:n Digital Universe -raportissa vuodelta 2011 ennustettiin datan määrän kasvavan vuosikymmenen loppuun mennessä noin viisikymmenkertaiseksi. Investointien tai tietojenkäsittelyn ammattilaisten määrä ei kasva samaa vauhtia. Vuonna 2013 YK:n raportissa “New Technology and the Prevention of Violence and Conflict” puolestaan todetaan, että vuonna 2012 ihmiskunta tuotti enemmän dataa kuin koko sitä edeltäneen historiansa aikana. Ylilyönti tai ei, tuon tyyppisiä toteamia on esiintynyt kiihtyvään tahtiin viime vuosina.

Vauhti

Määrän lisäksi haasteita syntyy datan tuottamisen kiihtyvistä tahdistista. Jo nyt on mahdotonta tallentaa pysyvästi kaikkea tuotettua dataa ja tulevaisuudessa tilanne tulee olemaan vielä pahempi. Datavirtoihin tulisi kuitenkin päästä käsiksi, jotta niitä voisi hyödyntää. Big datassa on yhtä lailla kysymys suurten datamassojen tallentamisesta kuin valtavien datavirtojen yhdistämisestä ja älykkästä hyödyntämisestä.

Variaatio

Kolmas V-kirjain tuo esille dataan ja sen rakenteeseen liittyvän haasteen. Data on yhä vaihtelevampaa ja datalähteiden määrä kasvaa kiihtyvästi. Standardoituihin formaatteihin nojaavat ratkaisut eivät kykene vastaamaan haasteeseen ja mitä suurempi osa datasta on vailla struktuuria tai strukturoitu tavalla, jota tietojärjestelmät eivät voi tehokkaasti hyödyntää, sitä suurempi haaste on menestyä dataintensiivisessä tulevaisuudessa.

Muita v-alkuisia ilmiöön liitettyjä määritteitä

Veracity, value, virality, validity, viscosity, viability ja myös muita v-alkuisia sanoja on käytetty kuvaamaan big data -ilmiön olemusta. Näistä kaksi tämän hetken suosituinta ehdokasta neljänneksi V-kirjaimiksi ovat veracity ja value. Veracity viittaa dataan oikeellisuuteen liittyvään epävarmuuteen ja value puolestaan sen arvoon käyttökontekstissaan. Olennaista näissä ei kuitenkaan ole se, mitä V-yhdistelmää käytetään, vaan se, että ilmiössä on kyse muustakin kuin datan suuresta määrästä.

Data ilmiön keskipisteessä

Big datan yhteydessä 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 mittakaavassa tai muodossa ole arvokasta, vaan vasta sen organisointi ja analysointi tekee siitä arvokasta.

1.1 Kansainväliset big data –markkinat

Big datan tarkka käsitteellinen rajaaminen on mahdotonta ja siksi tarkkaa arviota markkinoiden koosta ei voida antaa. Yksi tyypillinen ja paljon siteerattu arvio on IDC:n vuoden 2012 raportissa esitetty 9,7 mrd \$ vuonna 2013. Kasvuprosentiksi IDC arvioi vuositasolla noin 30 %.

Toimijalistauksissa pätee sama ongelma, mitä edellä mainittiin markkinoiden koon arvioinnista. Tässä aakkosjärjestyksessä esimerkinomainen otos tyypillisesti listauksissa esiintyvistä nimistä:

Amazon	Google	Microsoft	SAP
Cisco Systems	Hortonworks	NetApp	SAS
Cloudera	HP	Oracle	Splunk
Dell	IBM	Pivotal	Teradata
EMC	MapR	QlikTech	VMWare

Gartnerin tunnetun hypekuvaajan mukaan vuoden 2013 lopulla ollaan big datan kohdalla saavuttamassa ylimitoitettujen odotusten huippua. Matkaa tuottavuuden tasangolle on vielä, mutta taloustieteilijä Paul Krugmanin arvion mukaan on todennäköistä, että ilmiön mittasuhteita ja vaikutuksen suuruutta ei vielä nähdä ja että pitkällä aikavälillä on

odotettavissa mittakaavaltaan vielä merkittävämpiä muutoksia eri toimialojen kilpailun logiikassa ja julkishallinnon toiminnassa kuin mitä nyt arvioidaan. Tähän mahdollisuuteen kannattaa tarttua.

1.2 Big data -selvityksen lähtökohdat

Big data on tunnistettu merkittäväksi muutosvoimaksi ja myös mahdollisuudeksi. Suomen näkökulmasta se tarjoaa tilaisuuden hyödyntää jo olemassa olevia vahvuuksia ja osaamista kansallisen kilpailukyvyn parantamiseksi ja yksittäisten yritysten jalostamiseksi kansainvälisiksi menestystarinoiksi.

Big data liittyy läheisesti muihin Suomen kriittisiin avainosaamisalueisiin. Tietoliikenteen osajina olemme perinteisesti käsitelleet suuria datamääriä. Tietoturva on tärkeää kaikissa big data -tyyppisissä sovelluksissa ja tietovarannoista louhittavien tiedonjyvästen integrointi vaatii vahvaa ohjelmisto- ja tietojenkäsittelyosaamista. Julkisella puolella tietovarantojen avaaminen, yhteinen ICT-palveluarkkitehtuuri ja perinteisesti Suomessa hyvin toimiva julkinen-yksityinen-yhteistyö avaa mielenkiintoisia mahdollisuuksia.

Lähde: 21 polkua Kitkattomaan Suomeen, ICT 2015 -työryhmän raportti

Keväällä 2013 järjestettiin Liikenne- ja viestintäministeriön KIDE-ohjelman puitteissa big data –työpaja, jossa aihetta lähestyttiin useasta näkökulmasta paneelikeskusteluin ja lyhyin esityksin. Työpajan yhteydessä käyty keskustelu toimi innoittajana myös tämän selvityksen laatimiselle.

Se, osataanko big data –ilmiön tarjoamaan mahdollisuuteen tarttua ja mikä on lopputulos, riippuu paljon siitä, mitä päätöksiä nyt tehdään. Ilmiön mahdollisuudet ulottuvat kaikille sektoreille ja koskettavat kaiken kokoisia organisaatioita.

2. Tavoitteet ja menetelmät

Big data on ilmiönä nuori ja nopeasti muuttuva, joten suomalaisen big data -kentän kartoitus pyrittiin tekemään kohtuullisen nopeassa aikataulussa ja yleisellä tasolla. Selvitys suunniteltiin sellaiseksi, että vastaaminen on helppoa ja nopeaa, jotta otoksesta saadaan riittävän kattava. Selvitykseen sisältyvässä kyselyssä pyrittiin välttämään eksplisiittisiä listoja tällä hetkellä pinnalla olevista tuotteista ja teknologioista, jottei alan nopea kehittyminen vanhentaisi selvityksen tuloksia. Kysymysten asetelulla pyrittiin valottamaan big data -ilmiötä erityisesti osana yritysten, julkishallinnon yhteisöjen ja oppilaitosten kokonaistarjontaa. Itse kyselytutkimus sisälsi eri tahoille räätälöidyn Internet-kyselyn ja poimintoja eri roolien edustajien yksityiskohtaisemmista haastatteluista. Selvityksen loppuraporttiin sisällytettiin big data -ilmiön yleinen kuvaus ja kooste tulosten analyysistä.

2.1 Internet-kyselyt

Big data -selvitystä varten luotiin web-lomakkeet eri vastaajarooleille: konsultit, palveluntarjoajat, julkishallinto, oppilaitokset ja big dataa todennäköisimmin hyödyntävät yritykset (asiakkaat).

Lomakkeet rakentuivat yleisistä, kaikille yhteisistä kysymyksistä sekä roolikohtaisesta osuudesta. Lomakkeen kysymyksistä suurin osa oli vapaita tekstikenttiä, joiden avulla pyrittiin antamaan mahdollisuudet hyvin heterogeeniselle vastaajajoukolle esittää näkemyksiään vapaasti ja toisaalta vältettiin määrittelemästä tai rajaamasta big data -ilmiötä turhan tarkkaan. Yksiselitteisemmin määriteltyihin käsitteisiin perustuvat kysymykset rajattiin suuruusluokkiin tai valmiisiin valintalistoihin vastaamisen nopeuttamiseksi ja tulosjoukon yhdenmukaistamiseksi.

2.2 Haastattelut

Ivorio haastatteli selvityksen aikana eri rooleihin kuuluvista tahoista aktiivisesti kansallisella tasolla osallistuneita tahoja myös yksityiskohtaisemmin. Haastatteluissa annettiin osallistujien kuvata big dataan liittyviä toimintojaan sekä vapaasti että kyselyn rungon mukaisesti kategorisoituna.

2.3 Yksilöitävyys

Selvitykseen sisältyvä kysely rakennettiin siten, että sen voi täyttää yksilöimättä tarkalleen vastaavaa tahoja. Kyselyssä annettiin erikseen mahdollisuus antaa eksplisiittinen lupa käyttää vastaajatahon nimeä raportin osana. Lähtökohtana oli, että vastaajien tietoja ei yksilöidä tilastollisessa osuudessa, vaan yksilökohtaiset esimerkit nostetaan esiin analyysivaiheen tekstissä. Myös tarkemmissa haastatteluissa sovittiin erikseen luvista käyttää haastattelun sisältöä osana raporttia.

3. Big datan ja keskeisimpien teknologioiden esittely

Big datan alle sijoitetaan kirjava joukko tuotteita, palveluita ja teknologioita, joista osa on ollut olemassa jo ennen kuin big datasta ilmiönä alettiin puhuakaan. Näiden vakiintuneiden ratkaisujen soveltuvuutta suurten ja nopeasti kasvavien datamassojen haasteisiin vastaamiseen ei voida kiistää, koska ilmiön määritelmä on niin väljä. Tässä tarkastelemme kuitenkin pääasiassa uusia ratkaisuja, jotka pyrkivät joko käsittelemään ja tallentamaan suuria datamassoja tai hyödyntämään uudentyypistä analytiikkaa.

3.1 Hadoop

Tunnetuin yksittäinen uusi teknologia on Hadoop. Se on avoimen lähdekoodin Apache-lisensioitu ohjelmisto, joka julkaistiin vuonna 2006. Innottajina toimivat Googlen julkaisemat artikkelit tiedon tallentamisesta (GFS, Google File System) ja suurten datamassojen hajautetusta analysoimisesta (MapReduce). Hadoopin vastineet näille olivat nimeltään HDFS (Hadoop Distributed File System) ja MapReduce. Jälkimmäiseen Googlella on Yhdysvalloissa useita patentteja, joiden tarkoitus on myöntämismuutoksella vuonna 2010 annettujen lausuntojen mukaan suojata yritystä ja sen ydinteknologioita.

Hadoop sellaisenaan ei todennäköisesti ole kiinnostava ratkaisu monellekaan organisaatiolle. Raakadatasta organisaation toiminnan näkökulmasta hyödylliseen tuloksiin pääsemiseksi vaaditaan liikaa kehitystyötä, asetusten säätämistä, integraatioita ja testaamista. Todennäköisempi vaihtoehto Hadoopin hyödyntämiseksi on joko valmiin jakelun käyttöönotto tai Hadoopin käyttö osana laajempaa kokonaisuutta. Esimerkiksi Gartner ennusti vuoden 2013 alussa, että vuonna 2015 jo noin kaksi kolmasosaa kehittyneistä analytiikkaratkaisuista sisältää Hadoopin.

Yksinomaan Hadoopin ympärille liiketoimintansa rakentaneista yrityksistä tunnetuimmat ovat Cloudera, Hortonworks ja MapR. Omia Hadoop-jakeluitaan tarjoavat myös Amazon, IBM, Microsoft ja Pivotal. Pivotal on vuonna 2013 toimintansa aloittanut EMC:n ja VMWaren yhteinen yritys, jota rahoittaa myös General Electric. Hadoopiin tulevaisuuspotentiaalista suuntaaviivaa antaa vuonna 2008 perustetun startup -yritys Clouderan menestys. Yritys on saanut pääomasijoittajilta rahoitusta yli 100 milj.€ ja sen markkina-arvosta tehdyt arviot liikkuvat jo nyt yli miljardissa eurossa.

Suomessa Hadoopin mahdollisuuksiin on herätty myöhään ja osajapulasta on lähitulevaisuudessa tulossa merkittävä pullonkaula siihen nojaaville big data -hankkeille.

3.2 Muistinvarainen analytiikka

Toinen tunnettu big datan yhteydessä usein esille tuotu teknologia on muistinvarainen analytiikka. Siinä palvelinklusterin keskusmuistia käytetään datan tallentamiseen analyysin ajaksi, jolloin latenssi on mahdollisimman pieni ja analyysi nopeutuu huomattavasti verrattuna tilanteeseen, jossa tallennustilana käytettäisiin mekaanista kovalevyä tai SSD-levyä.

3.3 Pilvipalvelut

Pilvipalveluita tarjoavat yritykset ovat hanakasti tarttuneet big datan avaamiin mahdollisuuksiin. Näennäisesti rajaton tallennustila ja laskentakapasiteetti, jotka molemmat joustavat täysin tarpeen mukaan, sopivat hyvin ratkaisuksi moniin big datan esittämiin haasteisiin. Suurten julkisten datamassojen laaja levittäminen vaatii keskitettyä ratkaisua ja pilvipalvelut tarjoavat luonnollisen vaihtoehdon avointen ja

maksullisten datavarantojen keskitettyyn jakeluun siten, että ne ovat helposti yhdistettävissä dataa käyttävien organisaatioiden pilvipalveluissa säilyttämään omaan dataan.

Pilvipalvelut tarjoavat joustavuuden lisäksi myös hyvän alustan julkisten datavarantojen jakamiselle sekä yksityisten datavarantojen myymiselle yritysten ja muiden organisaatioiden käyttöön. Pilvipalveluun tallennettavalle datamäärälle ei ole kattoa ja koska yhä suurempi osa pilvipalveluiden tarjoajista tarjoaa myös analytiikkaa palveluna, voidaan suuriakin pilveen sijoitettuja datamassoja yhdistellä ja analysoida tehokkaasti. Tähän liittyen voidaan myös ennustaa, että tulevaisuudessa yhä suurempi osa datasta tullaan tallentamaan pilveen. Tätä puoltaa se tosiasia, että jos pilvipalveluista löytyy niin suuria datamassoja, että niiden siirtäminen ei ajallisesti tai kustannusten puolesta ole mielekäästä tai mahdollista, mutta niitä kuitenkin haluttaisiin käyttää, niin ei ole muuta mahdollisuutta kuin siirtää oma data samaan pilvipalveluun kuin missä tuo massiivinen datakin on. Siirtäminen puolestaan aikaansaa viivettä ja synnyttää kustannuksia, joten tämä luo painetta siirtää oma data pysyvästi pilveen, missä se on helposti yhdistettävissä samasta pilvipalvelusta löytyvien muiden datavarantojen kanssa.

Haasteena pilvipalveluiden kohdalla ovat tietoturva, -suoja ja juridiset kysymykset. Tunnetuimmat palveluntarjoajat eivät tarjoa mahdollisuutta rajata datan tallennuspaikaksi yksittäistä kansallisvaltiota, vaan rajausta tapahtuu EU-tasolla eli taataan datan olevan EU-alueella sijaitsevassa palvelinkeskuksessa. Näin toimivat esimerkiksi Amazon ja Microsoft. Tietoturva ja -suoja nousivat jälleen suuren huomion keskustelunaiheiksi pilvipalveluiden kohdalla kesällä 2013, kun Yhdysvaltojen viranomaisten ja suurten palveluntarjoajien tiedonvaihto nousi otsikoihin mm. Prism-hankkeen tullessa julki.

Suomella on tarjottavanaan palvelinkestusten sijaintipaikkana paljon vahvuuksia, kuten vakaa yhteiskunta, laadukas infrastruktuuri ja koulutettua työvoimaa. Esimerkiksi Google on rakentanutkin palvelinkeskuksen Haminaan ja Microsoft ilmoitti alkusyksystä 2013 invessoivansa noin 200 milj.€ omaan palvelinkestukseensa Suomessa.

4. Jo tunnistetut keskeiset tiedontarpeet ja kehityskohteet

Big data on ilmiönä Suomessa varsin nuori ja kypsytön. Kansainvälisestäkin ilmiö on vasta kasvuvaiheessa ja siirtymässä teoriasta käytäntöihin. Kuten kaikilla teknisillä ilmiöillä, myös big datalla on sekä puolestapuhujia että ankaria kriitikkoja. Ilmiön voidaan kuitenkin todeta jo ylittäneen kriittisen massan tärkeydessä siirtyäkseen pelkästä hypestä oikeasti merkittäväksi tietojenkäsittelyn alueeksi. Jo se, että kaikki suuret kansainväliset ohjelmistotalot ovat kiirehtineet ottamaan big dataan liittyvät tuotteet keskeiseksi osaksi portfolioitaan, kertoo ilmiöön kohdistuvista odotuksista. Kuten muillakin uusilla ilmiöillä, myös big datan osalla vallitsee epätasapaino tiedon janon ja tiedon tarjonnan välillä. Syväosaaminen on edelleen harvojen hallussa ja tarve tiedon laajempaan jakeluun on ilmeinen sekä saatavilla olevien hyötyjen valjastamiseksi että väärin käsitysten aiheuttaman kitkan poistamiseksi.

Big datan voidaan sanoa olevan tällä hetkellä pääosin Yhdysvaltojen ja Intian toimijoiden hallussa. Muualla maailmassa ollaan herätty mukaan valitettavan hitaasti. Ottaen huomioon big dataan kohdistuvat odotukset nimenomaan globaalien yritysten kilpailukyvyyn parantamisessa, on myös Euroopan maiden syytä kiirehtiä mukaan.

Keskeisiä tunnistettuja haasteita big datan kehityksessä ovat erityisesti tiedon jakamisen kanavien rajallisuus ja valtavien tietomäärien käsittelyyn sisältyvät oikeudelliset asiat.

4.1 Tiedon jakaminen

Big data -tietouden jakamisessa voidaan nähdä kaksi toisiinsa tiiviisti kytkeytyvää keskeistä ongelmaa: koulutuksen puute ja osaajien vähyys. Vaikka akateemisissa datatieteissä on Suomellakin pitkä historia, siirtymä big datan ymmärtämiseen ei ole sujunut riittävällä vauhdilla. Teoriassa muutokseen reagoiminen saattaa olla suhteellisen helppoa, mutta käytännön opetusohjelmien muutokset erityisesti keskeneräisten ja kehittyvien tieteenalojen osalta ovat osoittautuneet varsin vaikeiksi. Tietotekniikan osalta big data ei ole suinkaan ensimmäinen tällainen ilmiö. Valitettavasti ilmiön vakiintumista odotellessa menetetään usein se merkittävin kilpailuetu, joka rohkeammalla opetusohjelmien päivittämisellä voitaisiin saavuttaa.

Osaajapula on yhdistelmä koulutusmahdollisuuksien puutetta ja oma-aloitteisen osaamisen kehittämisen vähyyttä. Vaikka useat yritykset ovat havahtuneet big data -ilmiön kasvuun, ei monessakaan olla panostettu oman osaamisen kehittämiseen riittävästi. Yritysten business intelligence voi olla hyvinkin kypsää, mutta valitettavan usein siihen sitoutuneet henkilöt ovat niin työllistettyjä, ettei heitä ehditä jatkokouluttaa omaksumaan myös big data -murroksen tekniikoita ja mahdollisuuksia. Ne harvat, jotka tällä hetkellä ovat big dataan syvällisemmin perehtyneitä ovat työnantajilleen erityisen arvokkaita resursseja ja heitä tuskin nähdään jakamassa ymmärrystään esimerkiksi akateemisissa maailmassa. Oppilaitosten resurssit eivät useinkaan riitä kilpailemaan kaupallisten yritysten kanssa harvoista osaajista.

4.2 Datalähteiden tunnistaminen ja hyödyntäminen

Big data -ilmiö luo kasvavan paineen datalähteiden paremmalle ymmärtämiselle. Siinä missä aiemmin on tallennettu varsin hyvin tunnettua ja strukturoitua dataa omasta liiketoimintakoneistosta relaatiokantoihin ja sitä kautta tietovarasto/BI-käyttöön, ollaan nyt tuntemattomilla vesillä kaiken tämän ulkopuolisen datan osalta. On kysyttävä mitä yrityksen liiketoiminnalle merkityksellistä dataa jätetään keräämättä tai hyödyntämättä. On myös katsottava kauemmas ja tutkittava sekä kumppaniverkon tarjontaa että julkisia avoimia tietolähteitä. Ennen kaikkea on löydettävä näitä eri lähteistä ja eri muodoissa

tulevia datavirtoja yhdistäviä tekijöitä ja ymmärrettävä yhdistämisen kautta saavutettavia liiketoiminnallisesti kiinnostavia mahdollisuuksia. Tässä kokonaisuudessa tarvitaan vahvaa osaamista oman liiketoimintaympäristön ja liiketoimintaan keskeisesti liittyvän datan osalta sekä luovaa ajattelukykyä uusien näkökulmien havaitsemiseksi ja sitä kautta mahdollisten kilpailuetujen tunnistamiseksi.

Datalähteiden kirjon kasvaessa lisääntyy myös dataa tarjoavien tahojen ja niiden käyttämien teknologioiden valikoima. Uusia dataan liiketoimintansa perustavia yrityksiä syntyy ja valtiolliset rajat ja säännökset hämärtyvät pilvipalveluiden myötä. Vakaalla ja kestäväällä pohjalla toimivia yrityksiä voi olla vaikea erottaa opportunisteista ja jopa epärehellisin keinoin menestystä hakevista tahoista. Yksiselitteisiä luotettavuuden mittareita ei välttämättä edes ole olemassa. Samaan aikaan kaikkialla korostetaan datan olevan toimialasta riippumatta yksi yrityksen keskeisimmistä ja arvokkaimmista omaisuuseristä. Yhtälönä tämän aarreaitan avaaminen ja jakaminen suhteellisen uusien kumppanien kanssa ja tuntemattomien palveluiden avulla on vaikea päätöksentekijöille.

Monelle yritykselle nämä osaamisalueet ovat vaikeasti koottavissa samaan asiantuntijaan tai edes tiimiin. Asiantuntija-apua on varsin rajallisesti tarjolla, johtuen mm. big data -teknologioiden nuoresta iästä. Ilman riittävää asiantuntemusta big data -hankkeiden käynnistäminen ja edistäminen tuntuvat hankalalta ja riskialttiilta. Tämä kaikki luo merkittävän hidasteen koko ilmiön konkretisoitumiselle niin Suomessa kuin muuallakin.

4.3 Oikeudelliset asiat

Osaamisen ja ymmärryksen niukkuuden rinnalla suurin big data -ilmiötä jarruttava tekijä on juridisesti vaikeaselkoinen toimintakenttä. Tietojenkäsittelyssä on toki jo aiemmin jouduttu luomaan lainsäädäntöä ja yhteisesti hyväksytyjä tapoja tiedon käsittelyyn, jakamiseen ja julkaisemiseen, mutta big data avaa tätä kysymystä edelleen. Siinä missä aiemmin toimittiin pitkälti oman yrityksen tai palvelun tuottamien tietojen kanssa, lähdetään big datan myötä kokonaan uuteen ulottuvuuteen datan yhdistelyssä ja prosessoitavien volyymien kasvussa.

Kun tiedon tallennus- ja käsittelykapasiteetti nyt kasvaa räjähdysmäisesti, syntyy jatkuvasti uusia kysymyksiä erityisesti yksityisyyden suojan ja tekijänoikeuksien osalta. Yhdistelemällä mm. internetistä vapaasti ladattavia tietoja ja analysoimalla niitä kehittyneillä algoritmeilla saavutetaan helposti rajoja, joissa tieto muuttuu liian yksilöiväksi lainsäädännön nykyisissä puitteissa ja edellyttää esimerkiksi henkilörekistereihin sovellettavia ilmoitus- ja raportointikäytäntöjä. Tiedon lataaminen, tallentaminen ja yhdistely rikkoo myös helposti kansallisia tai kansainvälisiä tekijänoikeussäädöksiä. Oikeusturvan valvominen muodostuu todennäköisesti yhdeksi suurimmista haasteista tulevaisuuden tietoyhteiskunnassa.

Eryteisesti EU:lla tuntuu tällä hetkellä olevan huomattavia vaikeuksia tuottaa yhteisesti hyväksyttäviä pelisääntöjä datan käsittelyyn suurissa määrissä ja kansalliset rajat ylittäen. Tämä voi olla yksi merkittävimmistä tekijöistä EU:n hitaudessa kasvattaa kansainvälisesti kilpailukykyisiä big data -toimijoita.

5. Kyselyn roolien kuvaukset

Kyselyn tavoite on luoda kattava kokonaiskuva koko Suomen big data -kentästä. Tätä varten kyselyyn on haettu vastaajia vaihtelevista rooleista ja erikokoisista yrityksistä ja yhteisöistä. Osa kysymyksistä valittiin kaikille vastaajille sopiviksi, osa taas eriytettiin eri rooleja paremmin luotaaviksi. Raporttiin sisällytettiin sekä kaikille yhteisiä tilastoja että roolikohtaisia erotteluja. Roolikohtaiset kysymykset tiivistettiin keskeisten huomioiden kuvauksiksi ja osa vastauksista on nostettu lainauksiksi raportin tekstin yhteyteen.

5.1 Asiakkaat

Roolin asiakkaat alle sisällytettiin yrityksiä kaikilta toimialoilta. Yhteisenä nimittäjänä oli selvä potentiaali hyötyä merkittävistä datahankkeista. Osalla näistä yrityksistä oli laajaa, jopa globaalia liiketoimintaa, josta muodostuu valtavia datamääriä. Näistä esimerkkeinä suuret konepajayritykset, kuluttajaelektroniikkaa valmistavat yritykset ja perinteinen teollisuus. Osalla toiminta on transaktiointensiivistä, luoden myös suuria määriä dataa. Näistä esimerkkeinä teleoperaattorit ja finanssiyritykset. Suuret pörssiyritykset sisällytettiin lähes kaikki mukaan oletuksella, että myös niiden datamassat muodostuvat helposti suuriksi. Lisäksi otettiin mukaan yrityksiä, joiden toimialaan data liittyy keskeisemmin ja joiden liiketoiminta rakentuu vahvasti tiedon jalostamisen varaan. Näitä ovat tietotekniikan ja esimerkiksi peliteollisuuden yritykset. Asiakas-kategorian yritykset olivat kaikki yli 50 henkilön ja useiden miljoonien eurojen liikevaihdon kokoluokkaa.

5.2 Koulutus ja tutkimus

Koulutus ja tutkimus niputettiin yhden roolin alle. Tähän otettiin mukaan kaikki yliopistot ja ammattikorkeakoulut, joiden opetusohjelmiin sisältyy tietotekniikka, tiedonhallinta tai tietojenkäsittely. Etukäteen tiedettiin jo, että nykyisissä opetusohjelmissä big data on varsin vähäisessä roolissa, mutta oletuksena oli, että tämä globaalistikin valtavasti kiinnostusta herättänyt ilmiö vakiintuu osaksi tietojenkäsittelyn koulutusohjelmia jo lähitulevaisuudessa. Tutkimuspuolen tahoiksi valittiin sekä julkisrahoitteisia että yksityisiä tutkimuslaitoksia, lähes fokusalueesta riippumatta. Oletuksena oli, että nykyaikainen tutkimus synnyttää aina merkittäviä määriä dataa ja big datan mahdollisuudet ovat varmasti ajankohtaisia. Merkittäviä määriä tutkimusdataa luovat lääketieteen yritykset sisällytettiin kuitenkin jo asiakas-kategoriaan.

5.3 Julkishallinto

Julkishallinnon edustajia valittiin laajalti sekä ministeriöistä että niiden alaisista laitoksista. Mukaan pyydettiin erityisesti paljon henkilö- tai tapahtumatietoja käsitteleviä tahoja, joiden datamäärien voitiin arvioida nousevan helposti perinteisten tietokantaratkaisujen ylärajoille. Aktiivisia julkishallinnon vaikuttajia ja aiheesta kiinnostuneita oli alustavasti kartoitettu jo Liikenne- ja viestintäministeriön KIDE-ohjelman puitteissa keväällä 2013 järjestämän big data –työpajan yhteydessä. Myös ICT 2015 -raportissa big data ja avoin data listattiin merkittäviksi poluiksi suomalaisen kilpailukyvyn kehityksessä. Näiden taustojen perusteella haluttiin kartoittaa laajalti nimenomaan julkisrahoitteisten toimijoiden lähitulevaisuuden suunnitelmia.

5.4 Konsultit

Rooliin konsultit valittiin yrityksiä, joiden liiketoimintaan keskeisesti kuuluu tiedonhallinnan haasteiden ratkominen ja asiakasyritysten avustaminen tietojenkäsittelyn tarpeissa. Näitä ovat niin suuremmat monille sektoreille hajautuneet konsulttitalot kuin pienet, pilvipalveluiden tai big data ilmiön ympärille rakentuneet start-upitkin.

Konsulttiroolin yrityksiltä odotettiin yhdessä palveluntarjoajien kanssa vahvinta tietämystä ja selkeimpiä näkemyksiä big data -ilmiön nykytilasta ja tulevaisuudesta.

5.5 Palveluntarjoajat

Palveluntarjoajiin valittiin yritykset, joiden liiketoiminnasta merkittävä osa rakentuu nimenomaan tiedon käsittelyn ohjelmistoihin. Vaikka useilla näistä tahoista on myös konsulttitoimintaa omien tuoteportfolioidensa tueksi, haluttiin ne eriyttää konsulteista. Syväosaaminen big data -teknologioiden osalta on oletettavasti vahvinta juuri näissä yrityksissä. Suurin osa valituista vastaajista on kuitenkin suuria monikansallisia yrityksiä, joiden tuotekehitys voi olla kaukana Suomesta. Tämän oletettiin näkyvän myös vastauksissa.

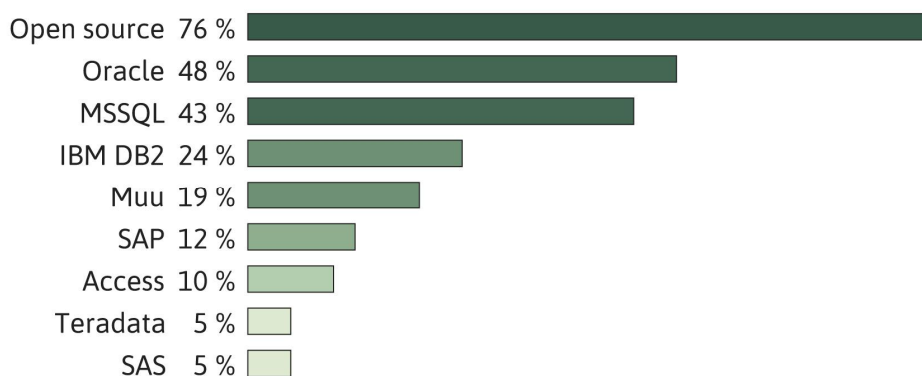
6. Selvityksen tulokset

6.1 Haastattelut

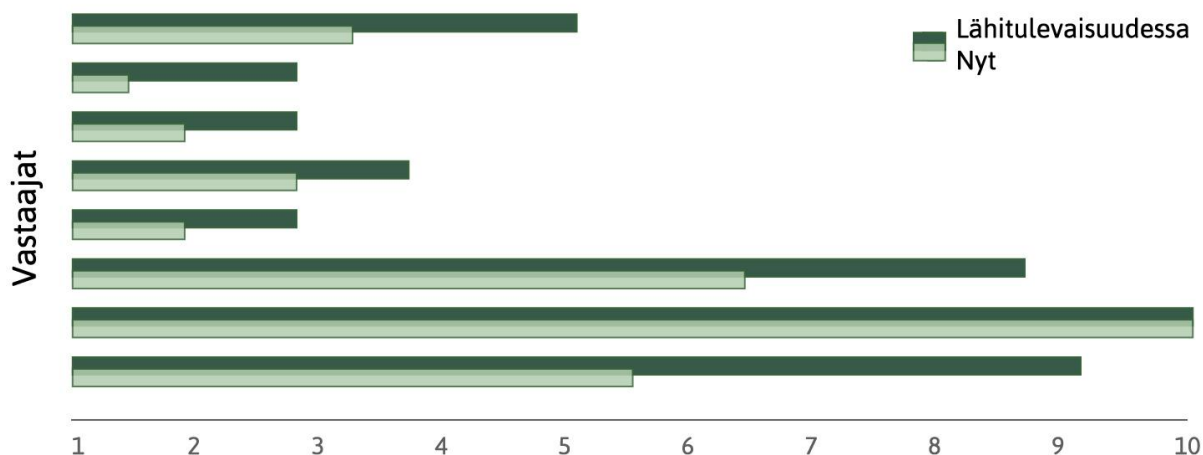
Haastatteluja suoritettiin kesän mittaan pääasiassa suurempien ohjelmisto- ja palvelutalojen sekä koulutus- ja tutkimuslaitosten keskuudessa. Kaikki haastateltavat osallistuivat myös täyttämällä omaa rooliaan vastaavan internet-kyselyn.

Big datan käytännön toteutuksissa ollaan monessa yrityksessä niin kriittisessä tai kilpailukyvyn kannalta keskeisessä vaiheessa, että yksityiskohtia ei haluta vielä paljastaa. Suurin osa kaupallisesta toteutuspotentiaalista on kuitenkin vielä hyödyntämättä ja projekteja korkeintaan vasta suunnitellaan. Monella yrityksellä suunnittelustakaan ei oikeastaan voida vielä puhua. Tätä havaintoa tukee myös suurten palveluntarjoajien vähäinen referenssien määrä. Osalla haastatelluista ei ollut Suomessa vielä yhtään kaupallista toteutusta, vaikka muuten tuotevalikoiman peitto mm. suuremmissa suomalaisissa yrityksissä olisikin erinomainen. Yhteistyön perinteiden odotetaan kuitenkin kantavan hedelmää lähitulevaisuudessa myös big data -hankkeiden ohjelmistopäätösten myötä.

Kaikilla merkittävillä kansainvälisillä ohjelmistotaloilla oli kuitenkin havaittavissa big datan merkityksen kasvu tuoteportfolioissa. Yrityssostot, tuotekehitys ja vanhojen tietovarastointi- ja analytiikkaohjelmistojen jatkokehittäminen ovat luoneet kokonaisia big data -ohjelmistoperheitä. Osaan on integroitu käyttöönottoa helpottamaan myös merkittäviä laitteistokomponentteja tai pilvipalveluja. Näiden suurten kokonaisuuksien varaan rakennetut käytännön esimerkit toimivat big datan yleisenä myyntipuheena lukemattomissa julkaisuissa. Asiakkaan näkökulmasta varjopuolena voidaan nähdä voimakas yhteen toimittajaan sitoutuminen, sillä vaihto ratkaisujen välillä ei ole tällä tasolla lainkaan triviaali tehtävä. Alla olevaan kuvaajaan koostimme esimerkkinä kyselyyn vastanneiden listaamat merkittävimmät nykyiset tallennusratkaisut suhteutettuna toisiinsa. Nähtäväksi jää, miten suuret pelurit pystyvät konvertoimaan tallennusratkaisujen asiakkaitaan myös big data -tuotteiden asiakkaita.



Kyselyyn vastanneiden ohjelmistotalojen arviot big datan merkityksestä nykyisessä ja lähitulevaisuuden liiketoiminnassa osoittivat ilmiön merkityksen kasvua kaikkien portfolioissa. Seuraava kuvaaja antaa suuntaa kasvunäkymistä.

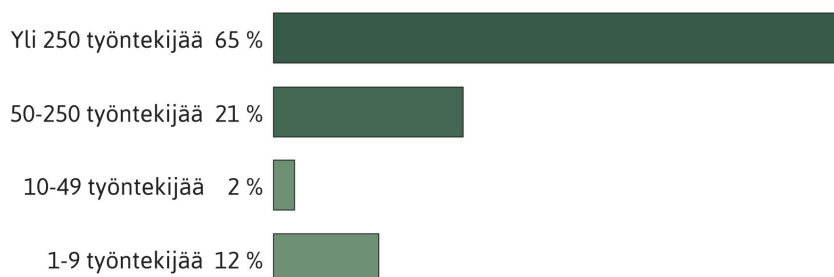


Henkilökohtaisesti haastatellut opetus- ja koulutussektorin toimijat olivat myös pitkälti samalla viivalla big datan integroimisessa osaksi tiedonhallinnan opetusohjelmiaan. Ilmiön taustalta on tunnistettu luontevia kasvupolkuja vahvasta tilastotieteen ja analytiikan osaamisesta sekä tietokanta- ja tietovarastointitietämyksestä. Big datan haasteina koetaan opetusmielessä yksiselitteisen määritelmän puute tai ilmiön rajaaminen sekä edelleen varsin varhaisessa kehitysvaiheessa olevien teknologisten ratkaisujen omaksuminen. Selvää kahtiajakoa oli havaittavissa tietoteknisesti painotettujen ja kaupallisesti painotettujen tahojen välillä. Tietotekniikka- ja matematiikkapohjalta arvioituna big data -ilmiössä nähtiin paljon harhaanjohtavaa markkinointipuhetta. Kaupallisesti painottuvien oppilaitosten suhtautuminen oli selkeästi avoimempaa ja hyväksyvämpää myös ilmiöön ladattujen valtaviin markkinaodotusten osalta. Siinä missä matemaattisemmalta pohjalta big data nähtiin jatkumona tiedonhallinnassa kasvavan tallennus- ja laskentakapasiteetin luonnollisena seurauksena, kaupallisemmat tahot uskoivat ilmiön luovan kokonaan uutta liiketoimintaa ja olevan merkittävä murros aikaisemmin hylätyn datan hyödyntämisessä ja merkittävien hiljaisten signaalien tunnistamisessa.

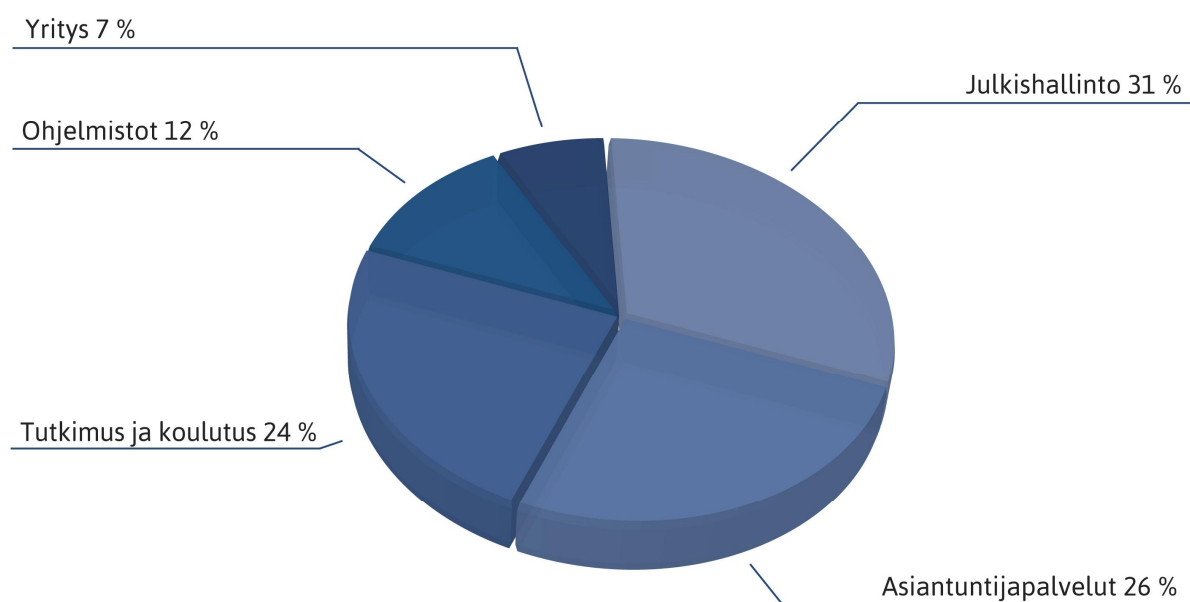
6.2 Verkkokysely

Verkkokysely toteutettiin välillä 28.6.2013 - 25.8.2013. Kesälomakausi rajoitti odotetusti vastausprosenttia, mutta muutaman muistutusviestin ja yhteyshenkilöiden henkilökohtaisen kontaktoinnin avulla saatiin varsin kattava kooste suomalaisesta big data -kentästä.

Kyselyn vastaanottajista ylivoimainen enemmistö oli jo etukäteen poimittu suuremmista toimijoista. Julkishallinnon instanssit, kansainväliset ohjelmistotalot ja pörssissä listatut yhtiöt muodostivat suuren osan osallistujista. Varsinaiset big datan erityisosajat puolestaan löytyivät pääosin pienistä ja keskisuurista ja varsin nuorista yrityksistä. Kyselyyn vastanneiden tahojen kokojakauma henkilöstön perusteella on havainnollistettu oheisessa kuvaajassa.



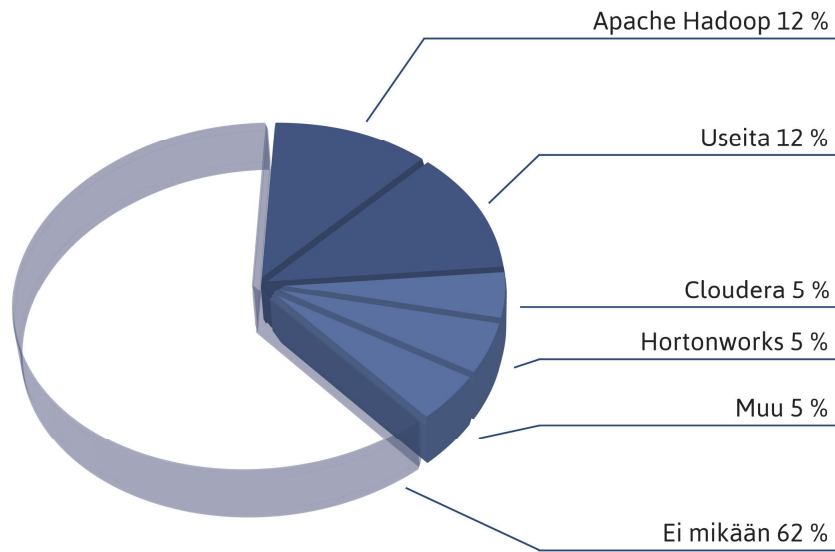
Kysely oli räätälöity eri rooleissa toimivia vastaajia varten sisältämään yleisten kysymysten lisäksi osan roolikohtaisia kysymyksiä. Roolit on kuvattu tarkemmin kappaleessa 5. Vastauksia saatiin tasaisesti kaikista rooleista. Vastaanottajalistalla edustajia oli kuitenkin selkeästi eniten koulutus- ja tutkimuslaitosten sekä julkishallinnon ja muiden suurien tietomääriä hyödyntävien asiakkaiden joukoista, mikä luonnollisesti painottui myös vastauksissa. Ohjelmistotarjonta on varsin rajoittunutta, keskittyen toistaiseksi suurten kansainvälisten toimijoiden joukkoon ja big dataan keskittyvä konsulttiosaaminenkin on toistaiseksi vielä harvojen tahojen hallussa. Lopullisten vastausten osalta jako rooleittain näkyy alla olevassa kuvaajassa.



6.2.1 Yleiset kysymykset

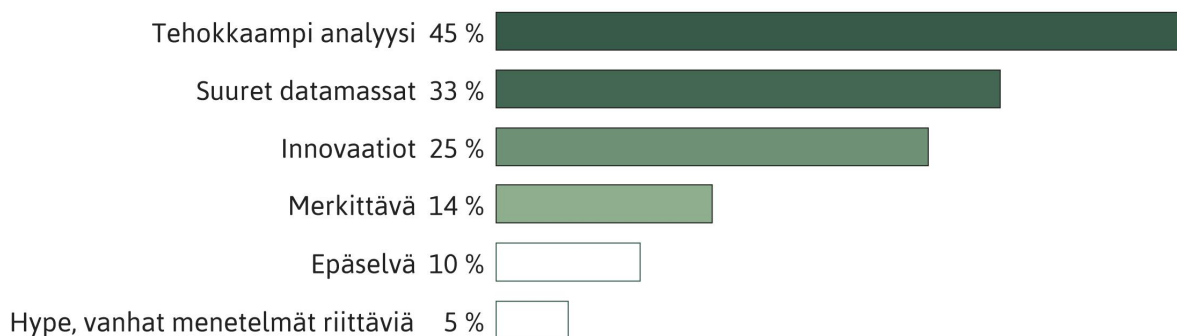
Yli puolet kyselyn kysymyksistä oli yhteisiä kaikissa roolissa toimiville vastaajille. Näiden perusteella voidaan arvioida big data -ilmiön jalkautumista Suomeen sekä konkreettisten käytännön lukujen että arvioidun nykyisen ja tulevan merkityksen kautta.

Kuten oletettua, käytännön kokemukset ovat vielä varsin rajallisia. Vastaajista yli puoilla (62%) ei toistaiseksi ole esimerkiksi kokemuksia Hadoopista, jonka voidaan katsoa olevan tämän hetken big datan teknisten toteutusten keskeisin elementti. Monilla teknistä osaamista tarjoavilla konsulttitaloilla kokemusta oli kokemusta useammastakin Hadoop-jakelusta Levinneisyyden osalta kaikki jakelut olivat tasaisesti edustettuina. Seuraava kuvaaja esittää kyselyn vastausten jakaumaa Hadoop-jakeluiden tämänhetkisestä hyödyntämisestä.



Kysyimme myös Hadoop-klusterien koosta. Vastaajien joukossa se vaihteli yhden koneen testikäyttäjistä muutamiin kymmeneen palvelimiin tuotantokäytössä. Kansainvälisessä mittakaavassa puhumme vielä siis näidenkin toteutusten osalta varsin maltillisesta kokoluokasta. Julkisia lähteitä vertailukohtana käyttäen, kymmenien palvelinten klusterit tuntuvat riittävän monille perinteisille suuryrityksille. Kokonaan omassa luokassaan toimivat puhtaasti internet- ja databisneksessä toimivat jättiläiset, kuten Facebook, eBay ja LinkedIn jopa muutamien tuhansien palvelinten kokoluokassa. Listan kärkisijalla on Hadoopin syntysija, Yahoo!, ällistytävällä yli 40 000 palvelimen klusterillaan. Tämän hetken suurin datayritys, Google, ei itse käytä ydinliiketoiminnassaan Hadoopia, mutta heidän big data -toteutuksensa lienee monella mittarilla maailman suurin.

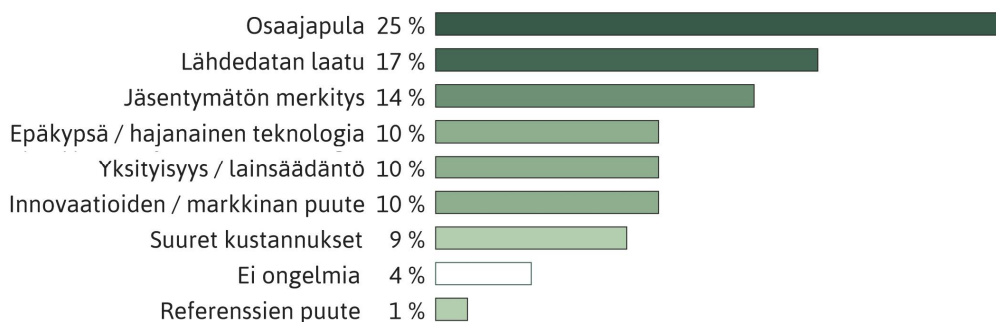
Kysyimme kaikilta vastaajilta big datan merkityksestä heidän organisaatiolleen, ilmiöön liittyvistä ongelmista ja tulevaisuuden visioista. Vastaukset vaihtelivat lyhyistä toteamuksista varsin moniulotteisiin kuvauksiin. Koostimme näistä tiettyjä teemoja, jotka toistuivat useissa vastauksissa. Alla olevissa kuvaajissa näkyvät eri teemojen väliset painotukset. Ensimmäisenä big datan merkitys.



Vahvimmin edustettuina olivat kaksi nimenomaan Hadoopin kaltaisten teknisten ratkaisuiden mukanaan tuomaa ominaisuutta: valtava tallennuskapasiteetti ja kyky prosessoida suuria datamääriä yksityiskohtaisemmin kuin koskaan. Tallennuskapasiteetti nähtiin monesti suoraviivaisesti kustannustekijänä ja nyt mm. pilvipalveluiden myötä laskevien kustannusten kautta voitiin todeta, että raakadataa kannattaa tallentaa enemmän kuin aiemmin. Prosessointiteho sinänsä ei ollut vastausten kärkipäässä, vaan nähtiin kaivattuna teknologisenä harppauksena kokonaan uudenlaisen analyysin tekemiselle. Hiljaiset signaalit, trendit ym. suuresta raakadatasta seulotut tiedot nähtiin

valtavana potentiaalina lähes joka toimialalla. Näiden perusrakenteiden varaan rakentuu luontevasti myös merkitykseltään kolmanneksi suurin mahdollisuus, eli dataan perustuvat uudet innovaatiot.

Big dataan liittyvät ongelmat kategorisoituivat myös varsin luontevasti alla olevan kuvaajan mukaisella jakaumalla.

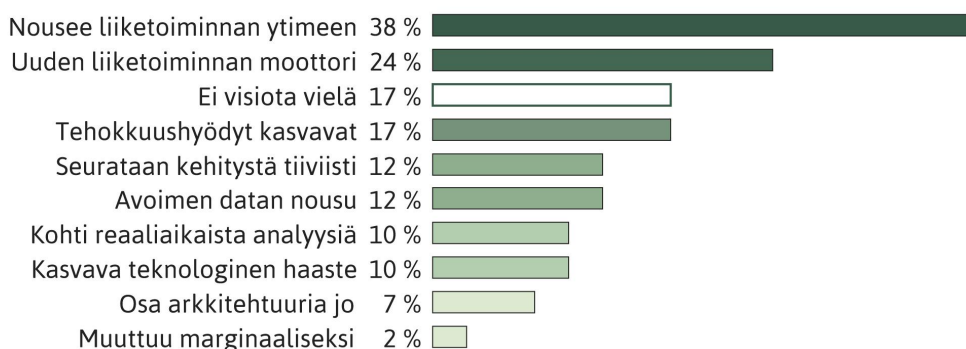


Kansainvälisestikin paljon pohdittu kasvava osaajapula oli odotetusti suurin ongelmakohta myös Suomessa. Teknisestä näkökulmasta teknologian kypsymättömyys ja lähdedatan kompleksisuus nähtiin myös merkittävänä kynnyksenä käyttöönotolle. Kolmantena haasteena etenemiselle nähtiin koko big data -ilmiön jäsentymättömyys. Ilmiö koetaan niin moniulotteisena, että esimerkiksi palveluiden tai asiantuntemuksen hankkiminen on vaikeaa. Tätä ongelmaa pahentaa entisestään pahasti perässä laahaava ja kansainvälisesti voimakkaasti vaihteleva lainsäädäntö.

Lainsäädäntö koettiin haastavana erityisesti yksityisyyskysymysten osalta. Julkisenkin datan käsittelyn nähtiin helposti muodostavan tahattomasti rekistereitä ja yksityisen datan yhdistämisessä julkisiin lähteisiin nähtiin tietovuotojen riskejä. Myös pilvipalveluina hankittavan tallennus- ja laskentakapasiteetin koettiin olevan ristiriidassa salassapidettävän aineiston käsittelyssä. Valtavin harppauksin kehittyvän teknologian myötä lainsäädäntö jää helposti vanhanaikaiseksi ja pahimmillaan muodostuu jopa innovaatioiden esteeksi.

Kysymys big data -visiosta tuotti vastaajien roolien erilaisuudesta johtuen varsin monenlaisia vastauksia. Yleistettynä oli kuitenkin helppo tunnistaa keskeiset trendit näidenkin joukosta.

Yli puolet vastaajista listasi big datan nousevan nykyisen tai tulevan liiketoiminnan keskeiseksi komponentiksi tai koki sen kasvattavan merkittävästi tehokkuutta. Muutamilla yrityksillä big data oli jo integroitu osaksi arkkitehtuuria ja visio sitä myötä myös varsin selvä. Myös epävarmuutta ilmiön merkityksen suhteen on olemassa, sillä lähes viidennes ei vielä osannut kuvata visiotaan tai ilmoitti seuraavansa toistaiseksi vain ilmiön kehitystä tiiviisti. Visiokysymyksen vastauksissa koettiin myös jonkin verran huolta teknologisen haasteen pahenemisesta datamäärien kasvaessa nopeammin kuin käsittelykapasiteetti ja sitä tukeva analysointiosaaminen. Seuraavassa kuvaajassa näkyy tulevaisuuden visioiden jakauma yksityiskohtaisemmin.



Julkishallinnon toimijoista useat korostivat big data -hankkeissaan nimenomaan avoimen datan mahdollisuuksia ja omaa rooliaan uuden liiketoiminnan, kehityksen ja tutkimuksen vauhdittajana. Tämä ilmiö on hienosti esillä myös monen julkishallinnon toimijan tavoitteissa ja saattaa kokemuksen karttuessa olla yksi merkittäviä vientivaltteja Suomelle lähitulevaisuudessa.

Vastaajien joukkoon mahtui myös kriittisesti big data -ilmiön merkitykseen suhtautuvia tahoja. Monet vastaajista eivät olleet vielä tutustuneet ilmiöön riittävän syvällisesti nähdäkseen siinä merkittävää potentiaalia. Toiset olivat tarkemman analyysin jälkeen todenneet nykyiset menetelmät riittäviksi vielä lähitulevaisuudessakin ja kokivat datamäärien kasvattamisen keinotekoiseksi omassa toiminnassaan.

Yksi selvityksen kannalta keskeisimpiä kysymyksiä oli Suomen asema tulevaisuuden big data -markkinassa. Tästä pyydettiin osalta vastaajista suoria arvioita, osa taas poimittiin tulevaisuuden visioista ja big datan merkityksen kuvauksista. Suomen mahdollisuuksien nähtiin jakautuvan aika tasaisesti tässä raportissa usein toistuvien teemojen välillä. Vahva perusinfrastruktuuri, ICT huippuosaaminen ja näiden tukema innovaatio loivat uskoa suomen mahdollisuuteen toimia tämänkin tietojenkäsittelyn muutoksen edelläkävijänä. Julkishallinnon aktiivisuus avoimen datan hankkeissa koettiin myös potentiaalisena keihäänkärkenä ja kansainvälistä kiinnostavuutta lisäävänä.

”Suomi on tunnettu avoimuudestaan ja läpinäkyvyydestään sekä erilaisista huippuosaajistaan, joten Suomella on paljon potentiaalia toimia eturintamassa sekä avoimen datan että big datan hyödyntämisessä.”

Helsingin kaupungin tietokeskus

”Innovatiiviset analytiikkaratkaisut perusteknologian päällä”

Joukkoon mahtui kuitenkin myös epäilyksiä Suomen kyvystä saavuttaa merkittävää roolia big datan kehityksessä johtuen mm. myöhäisestä reagoinnista ja investointikyvyn puutteesta. Alla olevassa kuvaajassa näkyy tarkempi jakauma näiden teemojen välillä.



6.2.2 Roolikohtaiset kysymykset: asiakkaat

Asiakkaat-rooli sisälsi yrityksiä, joiden toimialaan ja kokoluokkaan voi helposti kuvitella sisältyvän valtavia määriä tietoa. Toimialasta riippumatta kaikilla kyselyyn

vastaanottaneilla on varmasti myös tutkimatonta kehitys- ja liiketoimintapotentiaalia datamassojen laajemman ja syvemmän analyysin myötä.

Asiakas-roolin vastauksia saatiin mukaan kutsuttujen lukumäärään suhteutettuna ylivoimaisesti vähiten. Tämä on valitettavaa, mutta myös helposti ymmärrettävissä. Big data on ilmiönä niin ajankohtainen, että se ei ole voinut jäädä huomaamatta. Toisaalta, käytännön toimiin ei ole Suomessa juurikaan ryhdytty tai aloitetut hankkeet ovat erittäin todennäköisesti vielä salaisia tavoitellessaan merkittävää kilpailuetua. Nekin, jotka vastasivat, olivat selvästi lyhytsanaisempia kuin muiden roolien edustajat.

Vastaajista kukaan ei ilmoittanut omaavansa vielä minkäänlaista Hadoop-toteutusta tai tunnistanut organisaatiostaan Hadoop-osaajia. Muita tietokanta- ja tallennusratkaisuja ja niihin liittyvää kokemusta oli näillä suurilla yrityksillä luonnollisesti valtava kirjo.

Big datan konkretisoitumisessa käytännön kilpailueduksi on kuitenkin nähty potentiaalia. Vastauksissakin on tunnistettu sekä relaatiomallin rajoittuneisuus yhä kasvavien ja monimuotoisempien tietomäärien käsittelyssä että raakadatan tarkemmalla tarkastelulla löydettävät hiljaiset signaalit. Vastauksissa näkyy myös varovaisuus suhtautumisessa varsin nuoreen ja kypsymättömään murrokseen.

"Kaiken kaikkiaan datan ja siitä analytiikalla luotavan informaation merkitys tulee kasvamaan. Emme vielä ehkä oikein edes ymmärrä merkitystä."

Asiakas-roolin yritys

"Nyt opiskellaan ja selvitetään asiaa arkkitehtuurien kannalta potentiaalisena teknologiana"

Asiakas-roolin yritys

Ongelmiksi tunnistettiin vastaajien joukossa säännönmukaisesti osaajien vähyyys ja siitä johtuva epävarmuus toteutusten mahdollisuuksista ja kustannuksista. Osaamisen kehittämisessä nähtiin myös haasteita, koska luonteva polku perinteisestä data-osaajasta big data teknologioihin oli toistaiseksi epäselvä. Tulevaisuuden tiekarttaa oli kuitenkin maalailtu varsin realistisesti.

"Ratkotaan ensin akuutit ongelmat, joissa relaatiokannan rajat on jo vastassa. Sen myötä syntyneen osaamisen hyödyntäminen myöhemmin ratkaisemaan ongelmia, joita nyt ei edes osata unelmoida ja sitä myöden kilpailukyvyyn hakemista."

Asiakas-roolin yritys

Big datan arvioitiin olevan merkitykseltään tällä hetkellä korkea. Kehityskohteita oli tunnistettu sekä nykyisen datan tehokkaammassa hyödyntämisessä että uuden liiketoimintapotentiaalin tunnistamisessa. Erikoista vastauksissa oli edellä mainitut seikat huomioiden se, että IT-budjetoinnissa big data hankkeille oli varattu vasta 1-5%, ja suurin osa vastaajista ei uskonut osuuden kasvavan vielä vuonna 2014.

Vastaajat eivät ilmoittaneet olevansa aktiivisia big dataan liittyvän osaamisen hankkimisessa omaan henkilöstöön. Perusliiketoiminnan ollessa muualla kuin IT:ssä, tämä onkin varsin luonteva tilanne. Osaamistarpeen nähtiin kasvavan kaikkialla. Suomen historia tietotekniikan kärkimaana nähtiin tulevaisuudenkin mahdollisuutena, mutta myös huolta kansakuntamme aktiivisuudesta nähtiin.

"Suomi alkaa olla jäämässä jälkeen yleisesti ottaen datan hyödyntämisessä ja analytiikassa."

Asiakas-roolin yritys

Kaiken kaikkiaan suurten datamassojen omistajat ja siten potentiaalisimmat big data -ratkaisujen hyödyntäjät ovat vielä odottavalla kannalla. Ilmiön odotetaan jäsentyvän ja toimittajavalikoiman tiivistyvän. Oman alan menestyksekkäitä esimerkkejäkin kaivataan muilta ennen kuin omiin investointeihin halutaan ryhtyä. Nekin, jotka ovat konkreettisesti lähteneet toteuttamaan hankkeita, etenevät suhteellisen verkkaisesti.

6.2.3 Roolikohtaiset kysymykset: koulutus ja tutkimus

"Big datassa kiinnostavinta ei ole data, vaan siitä saatava hyöty. Hyödyn saamiseksi vaaditaan osaamista."

"BIG datan hyödyntäminen ei niinkään tarkoita SUURTA datamäärää vaan SUURIA (ja uusia) hyödyntämismahdollisuuksia."

Itä-Suomen yliopisto, tietojenkäsittelytieteen laitos

Koulutus- ja tutkimussektorin merkitys nykyisen ja varsinkin tulevan osaajapulan ratkaisijana sekä perustutkimuksen toteuttajana on suuri. Yksi keskeisimmistä ulkomaisissa selvityksissä tunnistetuista pullonkauloista uusien big data -teknologioiden hyödyntämisessä on osaajapula. McKinsey Global Institutin vuoden 2011 paljon siteeratussa raportissa "Big data: The next frontier for innovation, competition, and productivity", jossa todetaan pelkästään Yhdysvalloissa tarvittavan lähivuosina yli 100 000 datatieteilijää ja yli 1 000 000 datavallankumousta ymmärtävää johtajaa lisää. Globaalilla tasolla luvut ovat moninkertaiset ja tarve kasvaa nopeasti. Osaajapula voidaan tulkita menetetyiksi mahdollisuuksiksi:

"Big datan merkitys kasvaa jatkuvasti koska yhteiskunta digitalisoituu jatkuvasti. Paljon dataa jää analysoimatta koska osaaminen puuttuu."

Haaga-Helia Ammattikorkeakoulu

Maailmalta löytyy yksittäisistä yliopistoista jo runsaasti big data- ja kehittynyt analytiikka -nimikkeillä tarjottavia koulutusohjelmia, joiden kesto vaihtelee muutamasta päivästä puoleentoista vuoteen. Suomessa esimerkiksi Hadoopin käyttöä on esitelty jo vuosia esim. TKK:n yksittäisillä kursseilla, mutta big dataan liittyvää koulutustarvetta on aktiivisemmin herätty pohtimaan vasta viime vuosina. Puhtaasti big data -nimikkeen alla tai esim. Hadoopiin keskittyviä kursseja on vielä vähän, mutta niiden määrä lisääntyy nopeasti.

Osaamisongelman ratkaisemiseksi tarvitaan ajantasaista koulutusta yhdistettynä jo valmiiksi vahvaan osaamiseen.

"Tarve big dataan ja analytiikkaan liittyvälle tutkimukselle ja koulutukselle lisääntyy. Voimme vastata tähän olemassa olevalla osaamisella tiedonhallinnasta, järjestelmäkehityksestä ja analytiikasta, sekä kehittää ja tarjota uutta osaamista."

Jyväskylän yliopisto, tietojenkäsittelyn laitos

Olennaista koulutuksen suunnittelussa on tunnistaa nykyiset vahvat osa-alueet ja se, miltä osin niitä täytyy täydentää. Tilastomatematiikan, tietojenkäsittelyn, liiketoimintadatan analysoinnin ja ohjelmoinnin perusosaaminen on täydennettävissä big datan kannalta relevantiksi osaamiseksi, kun tunnistetaan puutteellisen osaamisen osa-alueet markkinoiden näkökulmasta. McKinseyn esille nostama kasvava pula suurten datamäärien tehokkaan hyödyntämisen problematiikkaa ymmärtävistä johtajista on haastava ratkaistavaksi. Syntyy tarve yhdistää tehokkaasti metodologinen ja tuloksia soveltava osaaminen.

"Näkökulma aiheeseen on tietotekninen. Emme ole kauppakorkeakoulu, joten big datan sovellukset liiketoiminnassa eivät kuulu keskeiseen profilliimme."

Helsingin yliopisto, tietojenkäsittelytieteen laitos

Datatieteilijän nimikkeellä kulkevan big dataan liitettävän uuden ammattikunnan osaaminen koostuu kokonaisuudesta, joka alkaa organisaation toiminnan ja menestyksen logiikan ymmärtämisestä, siihen liittyvien ongelmien tai haasteiden pohjalta muotoiltujen keskeisten kysymysten tunnistamisesta ja niihin vastausten löytämiseksi relevanttien datalähteiden löytämisestä. Kun datalähteet on tunnistettu, on seuraavana vuorossa datan analysointi eli raakadatasta hyödyllisen informaation, tiedon ja tietämyksen tuottaminen, levittäminen organisaatiossa sekä hyödyntäminen. Koko esitellyn ketjun täydellisiä taitajia ei olekaan, joten onnistuneeseen organisaation näkökulmasta keskeisten ongelmien ratkaisuun tähtäävään analytiikkaan tarvitaan tiimi, jonka jäsenten osaaminen täydentää toisiaan. Koulutuksen suunnittelun ja toteutuksen näkökulmasta tämä korostaa kokonaisnäkemysten tärkeyttä sekä koordinaatiota.

"Real experts, capable of utilising the technology and driving its development, are not easily available, although the demand for them is already high and rapidly increasing both in academia and industry. Collaboration and knowledge-exchange between big data experts and problem-driven science/industry/societal projects is lacking, but would be very beneficial."

"There is urgent need to train experts in database management, distributed/cloud computing, advanced data analysis methods, big data technologies, and theoretical background to maintain the competitiveness of Finnish researchers and companies in the global market. The most effective development and usage of big data methods often requires deep understanding of the field of their application, which is difficult to achieve for researchers individually. Interdisciplinarity and knowledge-exchange are important aspects of big data research."

Aalto-yliopisto, tietojenkäsittelytieteen laitos

Suuren haasteen lähitulevaisuuden koulutustarpeiden arvioinnille ja toteutusten suunnittelulle asettaa epäselvyys siitä, mikä osa big data -ilmiöstä on hypeä ja mikä taas todellista mullistusta. Määritelmällinen epäselvyys yhdistettynä nopeasti muuttuvaan ja kehittyvään teknologiaan tekee tästä vaikean haasteen vastattavaksi. On kuitenkin selvää jo nyt, että kyseessä ei ole puhdas hype ja että nykyisellä osaamisella ei täysin kyetä vastaamaan markkinoiden muuttuviin tarpeisiin.

"Within some research fields the size of the modeling projects in terms of variables and complexity increases roughly by an order of magnitude per year, which is much faster than the technological advances in general in ICT. This calls for entirely novel architecture for statistical inference in advanced hierarchical model, based on large-scale parallelization over thousands or tens of thousands of CPUs even for a single analysis. The future data and modeling tools have potential to enable huge advances e.g. in terms of human health concerning the risk of rapid spreading of virulent pathogens and multi-spectrum antibiotic resistance. Of course much of it is hype. But probably it will be something that will have a permanent impact on many areas of computing."

Helsinki Institute for Information Technology, HIIT

Kuten HIIT:n vastauksessa todetaan, sisältää big data ilmiönä suuria lupauksia. Pitkällä aikavälillä mullistus saattaa olla vielä suurempi kuin mitä nyt kyetään ymmärtämään. Osaamisen kehittäminen voi toimia merkittävänä katalyyttinä ilmiölle, joten siihen panostaminen on mielekästä. Suomen kansallisen kilpailukyvyn ja markkinoiden kehittymisen näkökulmasta huoli osaajapulasta tuli kattavasti ilmi vastauksista:

"Tulemme tarvitsemaan enemmän henkilöstöä parin vuoden sisällä. Tutkimus monilla alueilla on nyt jo siirtynyt Big Data alueeseen, mutta valitettavasti se on vielä tietotekniikkapainotteista. Osaajia, jotka osaavat sekä toimialakohtaisia ongelmia että big data -ohjelmointia, on liian vähän Suomessa."

Arcada-ammattikorkeakoulu

Osaamishaasteeseen vastaamiseksi on olemassa jo joitain kursseja ja opintokokonaisuuksia:

"Several basic and advanced courses related to big data are given at the department. A dedicated course "T-79.5308 Scalable Cloud Computing" directly addresses issues of big data. "

Aalto-yliopisto, tietojenkäsittelytieteen laitos

"Laitoksen antaman maisterikoulutuksen erikoistumislinjat 'Algoritmit ja koneoppiminen' ja 'Hajautetut järjestelmät ja tietoliikenne' kattavat yhdessä big data -alueen."

Helsingin yliopisto, tietojenkäsittelytieteen laitos

"Tarjoamme Big Data -kurssia valinnaisena edistyneille alemman korkeakoulututkinnon suorittajille ja ylemmän korkeakoulututkinnon suorittajille."

Haaga-Helia-ammattikorkeakoulu

Huomattava määrä lisäopetusta on suunnitteilla. Kuten uuden kurssiohjelmiston kanssa usein käy, on tarjonnan sisällön vertaaminen oppilaitosten välillä alkuvaiheessa vaikeaa. Seuraavia konkreettisia askeleita otetaan vastaajien lähitulevaisuuden tarjonnassa:

"Nimenomaisesti Big Dataan liittyvät kurssit ovat valmisteilla ja niitä käynnistellään tietojeni mukaan kuluvan lukuvuoden aikana."

Jyväskylän yliopisto, tietojenkäsittelytieteen laitos

"20-25 opintopisteen "Kehittynyt data-analytiikka johtamisessa" - koulutuskokonaisuus sekä maisteri- että kandidatin osaajille. Alkaa syksyllä 2013 ja jatkuu vuoden 2014 loppuun:

Kurssi 1. Kehittynyt data-analytiikka johtamisessa (orientoiva opintojakso)

Kurssi 2. Kehittynyt data-analytiikka - menetelmät

Kurssi 3. Big Data ja pilvipalvelut

Kurssi 4. Kehittynyt data-analytiikka - analyysitulosten tulkinta

Kurssi 5. Kehittynyt data-analytiikka - työvälineet

Lisäksi seminaarit sekä teemakoulutukset Pohjois-Savossa toimiville yrityksille ja organisaatioille."

I tä-Suomen yliopisto, tietojenkäsittelytieteen laitos

"Uusi data-analyysin maisteriohjelma käynnistymässä v. 2015."

Tampereen yliopisto, informaatiotieteiden tieteenalayksikkö

"Olemme hakemassa syksyllä luvat järjestää Informaatiotekniikan koulutusvastuuden alla Information Analytics erikoistumisopinnot. Harkitsemme myös englantinkielistä tekniikan master- ohjelmia."

Arcada-ammattikorkeakoulu

Lähitulevaisuudessa tilanne on siis muuttumassa nykyistä paremmaksi julkisrahoitteen koulutuksen osalta. Nähtäväksi jää riittävätkö panostukset ja ennen kaikkea osuvatko koulutusohjelmien sisällöt yksin markkinoiden osaajatarpeen kanssa. Akuuttiin osaajapulaan vastaamiseksi yliopisto- ja korkeakouluopetus ei välttämättä kykene

vastaamaan riittävän nopeasti. Jo työelämässä olevien ammattilaisten jatko- ja täydennyskouluttaminen nousee tässä avainasemaan. Painetta on sekä yksityisen koulutussektorin kasvuun että korkeakoulujen yhteydessä toimivien täydennyskoulutusohjelmien aktivoitumiseen. Ilmiöön liittyvän teknisen sanaston ja tuotekirjon epäkypsyys tuo vielä osaltaan oman vaikeutensa tuottaen laadukasta koulutusta. Tietyn tekniikan integrointi opetusohjelmiin voi osoittautua nopeasti virheeksi eikä muutoksia kyetä tekemään riittävän ketterästi. Tämä haaste on teknologiapainotteisessa kehityksessä varsin yleinen, mutta valmiita ratkaisumalleja on huonosti tarjolla.

6.2.4 Roolikohtaiset kysymykset: julkishallinto

Julkishallinnolla on monta roolia big data -markkinoilla. Yhtäältä se on mahdollistaja koulutus-, tutkimus- ja rahoittajaroolissa. Toisaalta lainsäädännön ja muun sääntelyn kautta se luo reunaehdot toiminnalle. Lisäksi julkishallinto on big dataan liittyvien palveluiden, tuotteiden ja teknologioiden käyttäjäasiakas. Avoimen datan näkökulmasta puolestaan julkishallinto on palveluntarjoaja eli se tarjoaa vastikkeetta markkinatoimijoiden käyttöön kasvavan määrän datalähteitä ja dataa. Suomessa ollaan avoimen datan hankkeissa kansainvälisesti vertailtuna pikällä. Datavarantoja on avattu ja rajapintoja pyritty yhdenmukaistamaan. Ympäristöviraston, Tilastokeskuksen, maanmittaus- ja ilmatieteenlaitoksen sekä Verohallinnon datavarantoja on jo avattu. Valtiontalouden kehyspäätöksen vuosille 2014-2017 mukaisesti datavarantojen avaamista jatketaan järjestelmällisesti myös tulevaisuudessa. Koko Euroopan Unionin sisämarkkinoiden näkökulmasta avoimen datan määrä lisääntyy merkittävästi, kun uuden direktiivin 2013/37/EU:n määrittelyt julkishallinnon datan julkistamisesta muuttuvat vuoden 2015 heinäkuuhun mennessä kansalliseksi lainsäädännöksi.

Directive 2003/98/EC should therefore be amended to lay down a clear obligation for Member States to make all documents re-usable unless access is restricted or excluded under national rules on access to documents and subject to the other exceptions laid down in this Directive. The amendments made by this Directive do not seek to define or to change access regimes in Member States, which remain their responsibility.

By 18 July 2015, Member States shall adopt and publish the laws, regulations and administrative provisions necessary to comply with this Directive.

Uuden liiketoiminnan vauhdittajan rooli muodostunee erittäin merkittäväksi suomalaisen datakeskeisen liiketoiminnan kehitykselle. CapGemini arvioi julkaisussaan The Open Data Economy, että koko EU-alueella avoimen datan suorat taloudelliset vaikutukset ovat noin 40 mrd euroa vuonna 2013. Datavarojen avoimuus on myös tunnustettu tietyillä toimijoilla keskeiseksi olemassaolon edellytykseksi.

"Tietokeskuksen yksi keskeisimpiä tehtäviä on avata pääkaupunkiseudun kaupunkien julkisia tietovarantoja ns. avoimena datana Helsinki Region Infoshare -palvelunsa (www.hri.fi) kautta kenen tahansa vapaasti ja maksutta hyödynnettäväksi."

Helsingin kaupungin tietokeskus

Useilla julkishallinnon toimijoilla oli jo käynnissä olevia tai lähiaikoina käynnistyviä, jo budjetoituja hankkeita. Kokoluokaltaan nämä olivat enimmäkseen suuria, yli miljoonan euron budjettiarvioilla, ja sellaisenaan jo merkittäviä työllistäjiä sekä kansallisen kokemuksen kartuttajia.

"Nyt: Tiedonluovutusratkaisun hankinta - tavoitteena vastata sekä avoimen datan kysynnän kasvuun ja varautua myös big-data ratkaisujen kehittämiseen. Suunnitteilla: Autojen paikka- ja kilometritiedon kerääminen ajoneuvorekisteriin ja näiden yhdistäminen mm. onnettomustilastoihin."

Trafi

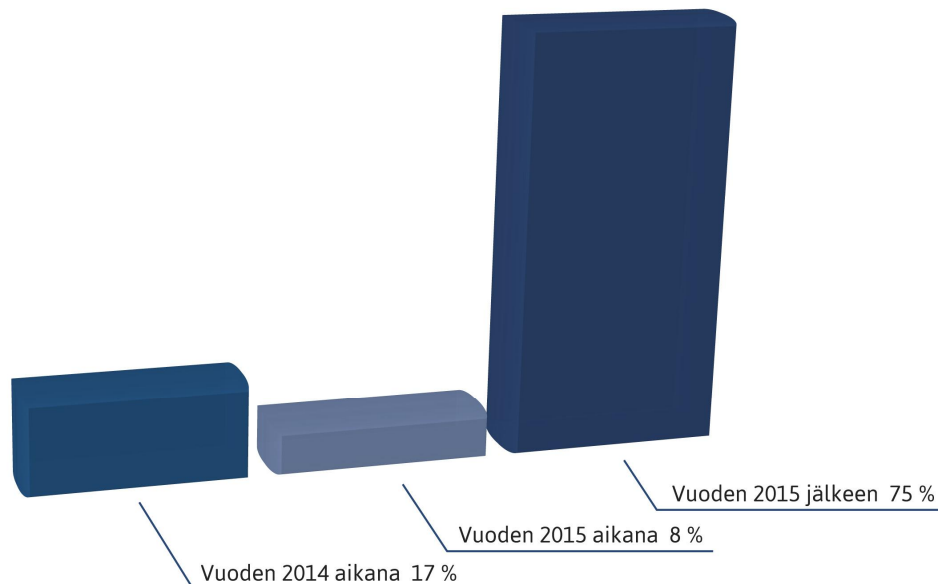
*"Nyt: Rakennus - ja Kiinteistötietojen uudistamishanke (RaKi)
Suunnitteilla: Henkilötietojen uudistamishanke (HeTi?)"*

Väestörekisterikeskus

Kysymykseen tärkeimmistä yritys yhteistyökumppaneista saatiin varsin vähän vastauksia. Yhteistyötä näytetään tehtävän enimmäkseen julkishallinnon yksiköiden kesken ja sisäisesti, sekä toimialan erikoisosaajien kanssa. Big data -tarjonnan lisääntyessä mm. perinteisissä suuremmissa konsulttiyrityksissä, on yhteistyön kasvu toki helposti nähtävissä. Tekniikan ja analytiikan kehityksessä yhteistyön rohkeus yli organisaatio- ja omistusrajojen voi parhaimmillaan olla merkittävä kiihdytin hankkeiden etenemiselle ja osaamisen kehitykselle. Juuri osaamisen ja innovatiivisuuden kautta big datan nähtiin avaavan Suomelle todennäköisimmän reitin kilpailukyvin kasvuun.

Huippuosaamisella on vaikutusta niin julkishallinnon sisäiseen toiminnan tehokkuuteen ja läpinäkyvyyteen kuin pysyvän kilpailuedun saavuttamiseen globaaleilla markkinoilla. Opetus- ja kulttuuriministeriö

Lopuksi kysyimme julkishallinnon toimijoilta arviota big datan arkipäiväistymisen ajankohdasta heidän toiminnassaan. Tässä ¾ vastaajista näki tuon tapahtuvan vasta vuoden 2015 jälkeen. Vähemmistö vastaajista koki ilmiön konkretisoituvan jo nopeammin. Alla kuvaaja jakaumasta:



6.2.5 Roolikohtaiset kysymykset: konsultit

Konsulttiyritysten vastaajat polarisoituivat kansainvälisesti toimiviin suuryrityksiin ja kotimaisiin pieniin erityisasiantuntemuksen yrityksiin ensimmäisten ollessa hienoinen enemmistö. Toimialoiksi listattiin ICT:n lisäksi vaihtelevasti omia erityisiä vahvuusalueita, kuten datatieteitä ja liiketoimintasektoreita.

Konsulttiyrityksissä valtaosalle Hadoop oli jo tuttu, pääosin Apachen perusversiona ja sen laajennuksina. Kaikilla toteutukset olivat kokoluokan perusteella kuitenkin vain testi- ja kehityskäytössä, ei siis osana palveluvalikoimaa esim. pilvipalveluina. Hadoop-toteutusten myötä osaamista on näihin yrityksiin jo kehittynyt, mutta toistaiseksi vasta harvoille asiantuntijoille. Suurimmat kansainväliset toimijat voivat toki tuoda muualta suurempiakin osaajajoukkoja tarvittaessa. Hadoop-analytiikan lisäksi useimmat konsulttiyritykset listasivat tarjoomaansa muita BI-työkaluja.

Konsulttiyritykset määrittivät säännöllisesti big datan jo johdannossa kuvatun kolmen V:n (volume, variety, velocity) kautta. Näiden yhteydessä korostettiin ilmiön konkretisoituvan hyvin monenlaisen osaamisen kautta: mm. analytiikan algoritmit, koostamisen ja yhteyksien tunnistamisen logiikka, reaaliaikakäsittely ja visualisointi.

Konsulttien keskuudessa big data nähtiin lähes yksimielisesti ensisijaisesti internet-liiketoiminnassa toimiville suurille tai kansainvälisille yrityksille välttämättömänä. Pienemmille yrityksille ja pääosin sisäisiin datalähteisiin keskittyville toimijoille big data nähtiin jopa tarpeettomana tai vähintään toisarvoisena, kun keskittyimistä korostettiin perinteisemmissä analytiikkaratkaisuisissa. Tässä päätelmässä on ehkä keskitytty nykyisen liiketoiminnan kehittämiseen enemmän kuin uusien ansaintamallien tai dataliiketoiminnan etsimiseen.

Big datan ongelmiksi konsulttiyritykset listasivat yleisimpien osaajapulan ja vakiintumattoman teknologian lisäksi datan säilytysvaatimukset mm. lainsäädännön puitteissa sekä helposti rikkoutuvan yksityisyydensuojan. Osaamiseen liittyvät haasteet pilkkottiin myös hienompiin osiin korostaen esimerkiksi algoritmien tehokkuuden optimointikykyä ja syvällistä monimuotoisen datan tulkintaosaamista. Osaamistarpeiden nähdään olevan myös niin laaja-alaisia, että ratkaisuja on etsittävä yhä useammin tiimeinä, joihin kuuluu sekä tietotekniikan, analytiikan että liiketoiminnan tuntijoita. Näiden haasteiden ratkaisemisen myötä big datan nähdään arkipäiväistyvän tulevaisuudessa. Suomella uskotaan olevan hyvät lähtökohdat ottaa tämän kehitysastekeen johtavan asiantuntijan rooli, mikäli vain uskalletaan siirtyä nopeasti sanoista tekoihin.

Puhtaasti big dataan erikoistuneita osaajia on konsulttiyrityksissäkin toistaiseksi vähän. Uskoa on kuitenkin, että kouluttautuminen, erityisesti teknologianäkökulmasta on mahdollista nopeastikin. Enemmän haasteita nähdään analytiikkoroolin täyttämässä. Kiinnostavan datan tunnistaminen eri liiketoiminta-alueilla ei ole välttämättä yhden ihmisen osaamisen tulosta eikä suoraviivaista teorian toteuttamista. Vastauksissa korostui edelleen monipuolisen osaamisen hakeminen tiimeissä.

"Collaboration of business and IT in traditional Business Intelligence engagements has been identified as one of the critical success factors a long time ago. However this will not be enough to cope with the challenges of harnessing Big Data. Close collaboration of various IT disciplines, especially in the areas of Enterprise Content Management, Data Management and Business Intelligence is required"

CGI

Suurimmat kyselyyn vastanneet konsulttiyritykset listasivat palveluihinsa lähes kaiken, mitä asiakasyritys voi big data -hanketta toteuttaessaan tarvita. Pienemmät paikalliset konsulttiyritykset keskittyvät lähinnä strategiseen liiketoimintaa tukevaan konsultointiin ja projektinjohtoon tai rajalliseen teknologiavalikoimaan. Yritykset eivät kuitenkaan big data -tarpeissaan voi olettaa kykenevänsä ulkoistamaan koko hankkeita, sillä keskeiseen rooliin tulee oman liiketoiminnan syvälinen analysointi. Siinä sekä käydään läpi vanhaa että tuotetaan uutta strategista pääomaa ja mukana täytyy olla yrityksen ydinryhmää. Merkittävässä hankkeissa voi helposti nähdä syntyvän myös haastavia tilanteita yritysten

strategisten linjausten suojelemiseksi, jos konsulttivoimia on mukana ja saman toimialan kilpailijoiden voidaan nähdä olevan kiinnostuneita vastaavasta palvelusta.

Teknologiapinossa konsulttiyritykset tarjoavat enemmän asiakaskohtaisuutta ja laajaa osaamista kuin erikoistumista tiettyihin ratkaisuihin. Tämä on luontevaa toistaiseksi, kun teknologiat big data -kentässä ovat vielä varsin tuoreita ja kehittyvät huimaa vauhtia. Erikoistumisen voidaan odottaa lisääntyvän ja yritysten fokusalueiden eriytyvän teknologian vakiintuessa.

6.2.6 Roolikohtaiset kysymykset: palveluntarjoajat

Palveluntarjoajat ovat keskeisessä roolissa big data -ilmiön sisällöllisessä täsmentymisessä ja dataan liittyvien haasteiden ratkaisemisessa. Big data -nimikkeen alla palveluitaan, tuotteitaan ja ratkaisujaan tarjoavien suurten globaalien toimijoiden määrä on kasvanut vauhdilla viimeisen kolmen vuoden aikana ja valikoimaa on sekä levennetty että syvennetty. Kyselyyn valitsimme edustavan otoksen niitä toimijoita, joilla on näkyvä edustus myös Suomessa ja joilla on selkeästi big data -ilmiön alle tunnistettuja ratkaisuja. Useilla big data on nimetty keskeiseksi liiketoiminnan kohdealueeksi lähitulevaisuudessa. Big data visiosta esitettyyn kysymykseen saimme vaihtelevia vastauksia, mutta useamman kuin yhden toimijan strateginen tavoite lienee lähellä tätä vastausta:

- Olla merkittävin Big Data -hallintaratkaisuja ja Big Data -liiketoimintasovelluksia tarjoava ohjelmistokumppani globaaleilla markkinoilla.

Vastauksissa ilmiön merkittävyys tunnistettiin ja se nähtiin pikemminkin strategisena muutosvoimana kuin operatiivisena hienosäätönä. Sitä ei myöskään nähty irrallisena aiemmasta kehityksestä tai tarjoomasta. Toimialanäkökulmasta big datan nähtiin ilmiönä koskettavan pitkällä aikavälillä kaikkia toimialoja.

"Big datan merkitys on suuri, koska se ikään kuin muuttaa käytäntöä siitä, kuinka muutamme dataa informaatioksi tai tietämykseksi."

Softability Group

"Microsoft näkee big datan yhdessä perinteisemmän BI:n kanssa osana laajempaa liiketoiminnan analytiikkaa. Big data -ratkaisujen tulee tukea integroidusti kaiken tiedon hallintaa, analysointia ja visualisointia."

Microsoft

"Big Datan hyödyntäminen ei ole toimialasidonnaista ja oikeastaan vain innovatiivisuus rajoittaa uusien palveluiden kehittelyä."

HP

Selvästi tunnistettiin kuitenkin se, että ilmiöön liittyy paljon uutta teknologiaa, uusia innovaatioita ja ennen kaikkea uusia osaamisvaatimuksia.

"Big Data -järjestelmien kehittäminen on aluillaan ja olemassa olevat järjestelmät ja teknologiat ovat vasta elinkaarensa alussa."

HP

Yleinen ongelma on tunnistaa ja löytää big datan mahdollisuudet, se miten aidosti hyödynnetään liiketoimintaa. Jos jokin on mahdollista tehdä ei vielä tarkoita että se olisi hyödyllistä. Toisaalta kaikkia mahdollisuuksia voi olla vaikeaa nähdä, jos organisaatiossa ei ole tehty asiaan liittyviä harjoituksia.

Rongo

"Suurimmat ongelmat liittyvät siihen miten liiketoiminta:

- *Saa oikeat järjestelmät ja työkalut kokoamaan informaation jota tarvitaan*
- *Antaa pääsyn johdolle heidän tarvitsemaansa tietoon ilman tukeutumista IT:hen*
- *Saa käyttöönsä järjestelmät jotka on suunniteltu toimialatarpeisiin*
- *Ymmärtää omistetun tiedon ja miten se käännetään liiketoiminnan parhaaksi*
- *Näkee tiedon oikeaan aikaan oikeassa muodossa."*

Oracle

Vaikka yleinen suhtautuminen ilmiöön heijasti pääasiassa näkemystä big datasta mahdollisuutena, ei meneillään olevaa datavallankumousta nähty täysin ongelmattomana. Muutoksen nopeus yhdistettynä uudenslaisiin osaamisvaatimuksiin tunnistettiin yhdeksi haasteeksi.

"Datan volyyymi, muutosvauhti ja sen sijainti hyvin erilaisissa tietovarannoissa tekee tiedon analysoinnin ja hyödyntämisen haasteelliseksi. Pelkän volyymin hallinta (Hadoop) ei riitä, vaan on kyettävä yhdistämään dataa myös perinteisistä tietovarannoista ja kyettävä analysoimaan ja visualisoimaan dataa tehokkaasti."

Microsoft

"Big Datan osaajatarpeet ovat vaativat. Vaatimuksena Big Data specialistille on ymmärtää liiketoiminnan logiikkaa, mitä tietoa liiketoiminta pääasiallisesti tuottaa ja kuinka tiedon laatua, tuottavuutta ja tehokkuutta voidaan parantaa. Jo kaikki tämä on laaja alue asiantuntemukselle, sillä liiketoiminnan logiikan ymmärtäminen vaatii todennäköisesti tietyn pohjakoulutuksen tai vähintään pitkäaikaisen toimimisen kyseisessä organisaatiossa."

Oracle

Lisäksi liiallinen painottuminen teknologiaan liiketoimintaan kytkeytyvän ymmärryksen sijaan nähtiin ongelmallisena.

"Lähestyminen tekniikan kautta ja tekniikan ehdoilla liiketoiminnan sijaan."

IBM

Käytännön toteutusten kohdalla, jossa teknologia on välttämättömyys, ongelmana puolestaan on vallitseva osaajapula. Uudet teknologiat mainittiin merkittävänä muutosvoimana ja mahdollistajana nopeasti kasvavien, monimuotoisten ja arvoa sisältävien datamassojen tulkinnessa. Useissa vastauksissa vuonna 2006 alkunsa saanut avoimen lähdekoodin ohjelmistoprojekti Hadoop nimettiin yhdeksi tärkeimmistä tai tärkeimmäksi tämän hetken yksittäisistä uusista teknologioista.

"Hadoop tulee olemaan big data järjestelmien ydinteknologiaa ja HP uskoo useimpien asiakkaiden käyttävän big data ratkaisuissaan sitä. HP:n strategiana on integroida omia tuotteitaan käytettäväksi yhdessä Hadoop-ratkaisujen kanssa."

HP

"Teknisiä osaajia ei juurikaan ole saatavilla tekniikan kohtuullisen nuoren iän vuoksi."

Softability Group

7. Selvityksen keskeiset havainnot

Suomen big data -kentän kartoitus vahvistaa entisestään käsitystä ilmiön jalkauttamisen keskeisistä haasteista. Toisaalta vastausten monipuolisuus ja vastauksista paistava innostuneisuus aihetta kohtaan antavat uskoa merkittävään kehitykseen tulevaisuudessa.

Haasteista eniten toisteltu on osaajapula. Tähän on lähdetty hakemaan ratkaisuja varsin aktiivisesti oppilaitoksissa, mutta näiden muutosten tuottamat hedelmät voidaan korjata vasta useiden vuosien päästä. Välittömään tarpeeseen ratkaisuja on etsittävä muualta, kuten täydennyskouluttamalla jo ennestään vahvoja tiedonhallinnan ammattilaisia uusiin teknologioihin ja muokkaamalla heterogeenisiä osaajatiimejä liiketoiminnan, analytiikan ja tietotekniikan taitajista. Kaupallisen koulutustarjonnan voidaan olettaa myös kasvavan, jolloin kohtuullisilla investoinneilla voidaan täydentää organisaatioiden osaamista kriittisten puutteiden osalta. Perinteisen koulutustarjonnan lisäksi on syytä tutkia myös erilaisia yhteistyömalleja, joissa osaamista ja kokemusta voidaan vaihtaa yhteisten hankkeiden ja tavoitteiden kautta esimerkiksi yritysten ja oppilaitosten välillä.

Suomen julkishallinnon toimijat ovat edistyksellisesti lähteneet toteuttamaan avoimen datan suunnitelmia. Kansalliset tietovarantomme voivat tarjota merkittävää lisävoimaa dataliiketoiminnan kasvuun ja sitä kautta kilpailukyvyyn ja viennin parantumiseen. Onnistuneilla avoimen datan hankkeilla voi olla jo sellaisenaan vientiarvoa ja toistettavuutta esimerkiksi naapurivaltioidemme etsiessä omia linjauksiaan datalähteiden avaamisessa. Julkishallinnon tukea tarvitaan kuitenkin edelleen, jopa kasvavissa määrin. Jotta dataliiketoimintaa ja uutta yrittäjyyttä voi syntyä, pitää myös datan käsittelyn pelisääntöjä selkiyttää. Oikeudelliset rajat pitää voida pukea riittävän kansantajuisiksi, jotta yritykset voivat rauhassa ja rohkeasti keskittyä innovoimaan omilla vahvuusalueillaan.

Big datan hyödyntäjien joukossa keskeinen tehtävä on edelleen oman strategisen tiekartan päivittäminen uuden ajan tiedonhallinnan mahdollisuuksia myötäileväksi. Tässä vaaditaan kykyä tunnistaa useita osa-alueita: uuden teknologian tarjoamia lisäkapasiteetin ja kustannustehokkuuden hyötyjä, oman datavarannon tarkempaa analyysiä ja hienompien signaalien tunnistusmahdollisuuksia sekä kokonaan uuden liiketoimintapotentialin löytämistä. Kaikilla osa-alueilla apua on tarjolla, vaikkakin rajallisesti. Näin ollen panostukset myös oman henkilöstön kouluttamiseen ovat varsin ajankohtaisia.

8. Jatkotoimenpide-ehdotukset

McKinseyn heinäkuussa 2013 julkaistussa "Game changers: Five opportunities for US growth and renewal" -raportissa listataan big data yhdeksi viidestä vakaata bruttokansantuotteen kasvua vauhdittavaksi tekijäksi tällä vuosikymmenellä. Tekstissä puhutaan Yhdysvalloista, mutta ajatukset ja perusteet ovat universaaleja ja sovellettavissa myös muualle. Suomen näkökulmasta big data ilmiönä tarkoittaa ennen kaikkea mahdollisuutta, johon kannattaa tarttua.

Syksyllä julkaistavassa Tulevaisuusselonteko 2030 -raportissa on big data yhdessä avoimen tiedon kanssa yksi viidestä kokeilusta, jotka käynnistetään syksyllä 2013. Marraskuun loppuun mennessä on tarkoitus tehdä ehdotus pitkän aikavälin ennakoinnin toimintamalliksi. Julkisen velan nopea kasvu ja huoli kestävyysvajeesta tekevät ajasta haasteellisen pitkän aikavälin visiointiin keskittymiselle, mutta jos ottaa rinnastukseksi 1990-luvun alun laman ja sieltä ponnistaneet menestystarinat, niin voi todeta, että vaikeinakin aikoina tulee olla rohkeutta siirtää katse myös tulevaisuuteen.

8.1 Koulutus

Big data -osaajista on huutava pula jo nyt. Nousevan Hadoop-teknologian taitajia on vähän ja sujuvasti pilvipalvelut ja big datan mahdollisuudet yhdistäviä osaajia vielä vähemmän. Koulutuksen vaikutukset tulevat markkinoille viiveellä, joten aikailuun ei ole varaa. Tarvitaan koordinoituja toimenpiteitä, joilla luodaan pohjaa tulevaisuuden ammattilaisten kehittymiselle ja kansallisen kilpailukyvyyn ylläpitämiselle big datan alueella.

8.2 Yritysten tukeminen

Puuttumatta terveen kilpailun reunaehtoihin on julkisen hallinnon mahdollista tukea yrityksiä, joiden strategian keihäänkärki nojaa big datan hyödyntämiseen. Tunnistettuja tukimuotoja ovat mm. seuraavat:

- perustutkimus
- aloittavien yritysten rahoitustarpeeseen vastaaminen
- kansainvälistymisessä tukeminen

Julkishallinnon lisäksi tässä on merkittävä mahdollisuus suurten yritysten ja riskirahoittajien tukemalle kasvulle. Kansainvälisesti ICT-alan suuryritykset toimivat hautomoina valtavalle määrälle uusia start-up yrityksiä. Suomessa vastaavaa toimintaa on enimmäkseen oppilaitosten yhteydessä, mutta uusia malleja on syytä etsiä täälläkin.

8.3 Julkishallinto kehityksen vauhdittajana

Julkishallinto itsessään voi hyötyä kiinnostavien datavarantojen ja -lähteiden tunnistamisesta ja hyödyntämisestä sekä tuottavuuden, kustannustehokkuuden että vaikuttavuuden alueilla. Nyt merkittävien julkisten palveluiden leikkausten aikakaudella kansalaisten palveluiden nykyaikaistuminen ja helpottuminen olisi myös varsin tervetullutta. Lisäksi datan saataville tuominen avoimen datan hengessä tuottaa dynaamisia kilpailua edistäviä vaikutuksia, jotka parantavat paikallisten yritysten kilpailukykyä, innovointivoimaa sekä pitkällä aikavälillä tuottaa lisää verotuloja työllisyyskehityksen ja yritysten taloudellisen voiton kasvun muodossa.

8.4 Yritysten innovaatiot ja yhteistyö

Kaupallisilla yrityksillä ei ole varaa jäädä odottelemaan big data -toimintaympäristön kehittymistä ja kypsymistä, vaan investointirohkeutta on löydettävä jo sitä ennen. Suurten yritysten ICT-budjeteista pienikin prosenttiosuus riittää teknologioiden koestamiseen. Eri osaamisalueita yhdistävien tiimien rakentaminen pitää integroida osaksi yrityskulttuuria. Innovatiivinen yhteistyö yritysten välillä, toimialasta ja kokoluokasta riippumatta, voi poikia uutta kasvua. Yritysten ja oppilaitosten yhteistyö esim. yrityshautomotoiminnassa voi nopeuttaa merkittävästi innovaatioiden nousua ideointivaiheesta käytännön liiketoiminnaksi.

8.5 Palveluntarjoajien lokaali aktivoituminen

Lähes poikkeuksetta kaikki julkiset big data -esimerkkitapaukset ovat ulkomailta. Näitä toistellaan kansainvälisten suuryritysten referensseinä kaikkialla. Ilmiön ja toteutusten yleistyessä on kuitenkin tärkeää pystyä viittaamaan onnistumisiin lähempänä kohdemarkkinaa, tässä tapauksessa Suomea. Kotimaisten, jopa saman toimialan onnistumisten kautta viestin välittäminen tehostuu ja edesauttaa uusien hankkeiden käynnistämisessä. Yleistä ymmärrystä tehostavia esimerkkitoteutuksia voisi harkita jopa yleishyödyllisinä hankkeina, jolloin arvo olisi mitattavissa enemmän PR- kuin tuloskehityksenä. Aktiivinen ja avoin osallistuminen myös koulutussektorilla luo vahvaa kuvaa yrityksen edelläkävijyydestä, avoimuudesta ja ajankohtaisuudesta. Teknologiapainotteinen yhteistyö oppilaitosten kurssikokonaisuuksien osana voisi myös tuoda helpotusta akuuttiin osaajapulaan.

9. Suomen data -foorumi

Selvityksen myötä on käynyt selväksi, että kiinnostus big dataa kohtaan on valtava ja halu lähteä korjaamaan uusien tiedonkäsittelyn keinojen hedelmiä on suuri. Big datan mahdollisuuksien haltuunotto koetaan kuitenkin läpi sektorien haastavaksi ja riittävää osaamista tai ymmärrystä ei tunnu olevan saatavilla. Big data on myös suurelta osin yritysten ja yhteisöjen välistä bisnestä, siirtyen perinteisestä oman datan hyödyntämisestä datamarkkinoihin. Tässä tarvitaan uusia toimintatapoja ja menetelmiä yhteistyön vauhdittamiseksi ja tiedonjaon parantamiseksi.

Monella taholla ollaan aktivoitunut big data -tietämyksen kehittämiseksi. Koulutusohjelmia syntyy, yritykset uudelleen kouluttavat datatieteilijöitä ja konsulttikenttäkin laajenee jatkuvasti. Asiantuntemus on kuitenkin vielä pitkän aikaa hajallaan eri yrityksissä ja yhteisöissä ja huippuosajista tulee olemaan pula vielä lähivuosina.

Yhteistyön tiivistämiseksi ja suomalaisen dataliiketoiminnan vauhdittamiseksi on nyt luotu uusi alusta: Suomen data -verkkofoorumi. Foorumin tarkoituksena on vauhdittaa tiedon kulkua niin julkishallinnon datahankkeiden kuin kaupallisten toimijoiden vauhdilla päivittyvän tarjonnan osalta. Foorumin ensisijainen tehtävä on yhdistää hajallaan oleva ja rajallinen suomalainen osaaminen datakeskeisen liiketoiminnan kehittyvien trendien (big data, avoin data) puitteissa. Foorumi tarjoaa mahdollisuuden kertoa alkavista hankkeista, välttää päällekkäisyyttä, kysyä ja antaa neuvoja sekä löytää kumppanuus ja synergiaetuja.

Suomen data -foorumin keskeisin arvo on avoimuus. Kaupalliset arvot syntyvät sivutuotteena yhteistyön kiteytyessä. Foorumi toimii erinomaisena kanavana sekä palveluita tarvitseville että tarjoaville tahoille. Markkinointiponnistelut foorumilla ovat asiantuntemuksen osoittamista ja sitä kautta syntyviä, eivät mainostamista. Foorumin säännöissä määritellään toimintamallit yksityiskohtaisemmin.

10. Loppusanat

Toivomme selvityksen tulosten herättävän ajatuksia ja aikaansaavan tahtotilan, jolla big datan mahdollisuuksiin tartutaan. Suomen data -foorumi tarjoaa kohtaamispaikan keskustelun jatkamiselle, ajatusten vaihtamiselle ja konsensuksen hakemiselle. Koulutustoimenpiteiden koordinointi, julkisten hankkeiden päällekkäisyyden välttäminen sekä eri kokoisten yksityisen sektorin toimijoiden kohtaaminen luovat yhdessä mahdollisuuden olla kansainvälisesti eturintamassa datavallankumouksessa.

Suuret kiitokset kaikille selvitykseen osallistuneille ja sille erinomaiset puitteet luoneelle liikenne- ja viestintäministeriölle. Tämä selvitys on vasta alkusoittoa, Suomella on erinomaiset mahdollisuudet johtavaan asemaan big datan kehityksessä. Edelläkävijän rooli on tarjolla esimerkiksi kansallisen avoimen datan mallimaana. Resurssikilvassa emme pienenä valtiona pärjää, mutta rohkean yhteistyön ja innovaation kautta voimme lunastaa paikkamme. Foorumilla tavataan!