

# **REAKTOR INNOVATIONS OY:N VASTAUS LAUSUNTOPYYNTÖÖN LUONNOKSESTA VALTIONEUVOSTON PERIAATEPÄÄTÖKSEKSI DATAN HYÖDYNTÄMISESTÄ LIIKETOIMINNASSA – MASSADATAN JA MYDATAN STRATEGISET LINJAUKSET JA TOIMET**

Ohessa Reaktor Innovations oy:n vastaus 28. tammikuuta saapuneeseen lausuntopyyntöön liittyen valtioneuvoston periaatepäätösluonnosta datan hyödyntämisestä liiketoiminnassa. Lausuntopyynnössä toivottiin myös ehdotuksia toimien ja toteuttavien tahojen täydentämiseksi.

## **Avoimen tiedon kulttuuri ja kokeilujen tukeminen**

Alan käsitteistö, esimerkiksi “massadata” ja “datatieteilijät” johtavat painotuksen helposti datan keräämiseen tai jakamiseen. Toinen, datatiede, johtaa näkökulma analytiikkaan liittyvien viimeisimpien teknisten menetelmien hallintaan. Datan keräämisen ja jakamisen lisäksi tarvitaan kokeita ja aktiivista datan tuottamista: toimenpiteiden seurauksista ei saa tietoa, jos toimenpiteitä ei kokeilla. Massadatan aikakaudella perinteisen tieteellisen kontrolloidun kokeen ideaa voidaan laajentaa koneoppimisen suuntaan, personoidun lääketieteen tapaan.

Datavetoisen (*data-driven*) ajattelun olennaisimmat osat liittyvät kulttuuriin, jossa painotetaan liikeidean, liiketoimintaan liittyvän ongelman, yhteiskunnallisen tai minkä tahansa ratkaistavan asian muotoilua sellaiseksi, että se on testattavissa ja joko olemassaolevaan dataan tai kärkeäkövään dataan nojautuen ratkaistavissa.

Uusien innovatiivisten kokeilujen tuottaminen onnistuu parhaiten verkostoissa, joissa viranomaiset, massadatan analysointiin erikoistuneet datatieteilijät, kuluttajamarkkinoilla toimivat yritykset ja erilaiset kolmannen sektorin organisaatiot toimivat yhdessä täydentäen toistensa osaamisalueita, data-ainestoja ja analytiikkaresursseja.

Viranomaistahot ylläpitävät kansallisesti ja taloudellisesti merkittäviä avoimia tietolähteitä. Esimerkiksi ilmakehän tilasta, maantieteellisestä demografiasta, ekonomisesta tilanteesta ja

hyödykkeiden hintatiedoista on saatavana avointa dataa. Avoimet aineistot mahdollistavat nopeat kokeilut ja demot.

## **Linjaus 1.1 Tietoaineistojen laajempaa käyttöä edistetään kannustavalla sääntelyllä**

Julkinen sektori voi parhaiten edistää kokeilukulttuuria tuottamalla dataa avoimesti kaikkien saataville erilaisten ohjelmistorajapintojen kautta. Reaktor pitää esitettyjä toimenpiteitä datan jakamiseksi oikean suuntaisina ja kannatettavina. Näiden toimenpiteiden lisäksi olisi syytä löytää keinoja joilla avoin data ja julkisen ohjelmistorajapinnat saataisiin osaksi kaikkien julkisten tietojärjestelmähankkeiden hankintaprosessia niin kunnissa, valtionhallinnossa kuin muissa julkisissa organisaatioissa.

Julkisissa hankinnoissa tulisi myös pyrkiä mahdollistamaan nopeat kokeilut soveltamalla hankintalain mahdollistamia kevyempiä hankintamalleja, kuten tavanomaista neuvottelumenettelyjä väljempiä neuvotteluja ja uutta innovaatiokumppanuusmallia.

Uusien avointen rajapintojen tuottamisen lisäksi tulisi pyrkiä luomaan nykyistä syvempää yhteistyötä ja vuoropuhelua dataa tuottavien julkisten ja yksityisten organisaatioiden, sekä dataa hyödyntävien kehittäjien välille. Pelkkä datan jakaminen tai antaminen julkiseksi ei takaa parasta mahdollista lopputulosta, vaan lisäksi tarvitaan datan jakajan toiminnan sekä dataa tuottavien ja hyödyntävien prosessien ymmärtämistä eli datan kontekstin laajempaa hahmottamista. Siksi esimerkiksi hackathon-tyyppisten tilaisuuksien lisäksi tarvitaan tiiviimpää molemminpuolista yhteistyötä ja kommunikaatiota datan omistavan ja siitä palveluja kehittävän tahon välillä.

Yksilöitä koskevan tiedon tallentamisessa ja hyödyntämisessä oleellista on on luoda luottamus sille, että tietoa käsitellään, jaetaan ja hyödynnetään vain yksilön itsensä sallimilla tavoilla ja ainoastaan osapuolille, jotka yksilö on valtuuttanut. Luottamuksen lisäksi on tärkeää, että oman tiedon jakamisesta ja avaamisesta koituu yksilölle selkeää lisäarvoa joka motivoi tiedon jakamiseen.

Kansalaisten luottamus siihen, ettei heitä koskeva data ilman heidän lupaansa kulkeudu tahoille, joita he eivät ole valtuuttaneet, on ensiarvoisen tärkeää. Tämä tulee ottaa huomioon myös uutta, niin sanottua verkkotiedustelulakia, valmisteltaessa. Lakiuudistuksen valmistelun aikana huolensa uuden lainsäädännön mahdollisesti mahdollistavasta "massaurkinnasta" ovat ilmaisseet niin useat akateemiset asiantuntijat kuin suomalainen tietoturvaohjelmistoteollisuus.

## **Linjaus 1.2: Kyvykkyyttä kasvattavalla datapolitiikalla mahdollistetaan kasvu ja vahvistetaan päätöksenteon tietopohjaa**

Kehittyvän ja uusia innovaatioita tuottavan datatalouden perusedellytyksiä ovat riittävät ohjelmointitaidot ja tilastotieteen ja -matematiikan ymmärrys. Datatieteen erityisosaajien lisäksi tietohallinto-organisaatioissa ja yrityksissä laajemminkin tarvitaan tilastotieteen perusteet osaavia työntekijöitä, jotka ymmärtävät datatieteen avulla saatuja tuloksia, osaavat hyödyntää niitä yrityksen liiketoimintaa ja strategiaa kehitettäessä, ja kykenevät arvioimaan niiden tilastollista luotettavuutta.

Pohja tälle elinkeinoelämän tarpeisiin riittävälle osaamiselle luodaan perus- ja toisen asteen opetuksessa. Tämän vuoksi ymmärrys tietokoneiden toiminnan periaatteista ja ohjelmoinnin perustaidot tulee opettaa jo peruskoulussa. Lukiossa tilastomatematiikan opetuksessa tulee hyödyntää sähköisiä aineistoja ja automaattista tietojenkäsittelyä. Korkeakouluissa eri tieteenaloilla opiskelijoille tulee opettaa tilastolliset perusmenetelmät ja kannustaa nykyistä enemmän määrällisen tutkimuksen tekemiseen laadullisen ohella esimerkiksi pro gradu -tutkielmissa ja muissa opinnäytetöissä.

Läheinen yhteistyö alan koulutusta antavien korkeakoulujen ja yritysten välillä on avain osaavien asiantuntijoiden kouluttamiseksi alan yrityksiin. Kun korkeakoulujen teoriaosaaminen ja käytännön työelämän kokemus yhdistyvät, saadaan kaikkien kannalta paras lopputulos.

Kun suunnitellaan koulutuspolkuja erilaisia datakyvykkyyksiä tukeviin tehtäviin, täytyy keskittyä myös analytiikkaan ja dataan liittyviin säädöksiin ja siihen mitä ne tosiasiallisesti tarkoittavat. Tietosuojaan ja henkilötietoihin liittyvän lainsäädännön tuntemus tulisi olla hyvää kaikissa dataa hyödyntävissä tehtävissä ja niiden perusteiden tulisi kuulua kansalaistaitoihin. Ainoastaan

johdon ja juridiikan asiantuntijoiden näkemys ja ei riitä sillä päätöksiä datan käytöstä tehdään myös teknisellä operatiivisella tasolla. Puutteellinen ymmärrys tietosuojalainsäädännöstä voi johtaa:

- datan hyödyntämisessä teknisesti ja organisatorisesti epätarkoituksenmukaisiin ratkaisuihin (esim. pilvipalvelun käyttämättä jättämiseen)
- siihen että dataa ensikädessä hyödyntävät eivät osaa vaatia perusteita ja reagoida oikein silloin, kun heille asetetaan tietosuoja rikkovia tavoitteita tai tehtäviä
- välinpitämättömyyteen toteuttaa kansalaisten tietopyyntöjä niin yksityisistä kuin julkisista lähteistä

## **Linjaus 2.1 Tuetaan datan hyödyntämiseen perustuvien toimintamallien kehittymistä ja markkinoiden avautumista**

Julkisen ja yksityisen sektorin yhteistyötä ennakoivan terveydenhuollon alueella tulisi edistää. Yhdistämällä yhteistyössä laitteita valmistavien yritysten kanssa esimerkiksi erilaisista puettavista laitteista, kuten askel- ja sykemittareista, saatavaa dataa julkisten organisaatioiden potilastietojärjestelmissä olevaan aineistoon pystytään luomaan uusia ennustavia malleja. Näillä pystytään tunnistamaan riskiryhmiin kuuluvia potilaita ja ehkäisemään sairauksien syntymistä.

## **Linjaus 2.2 Kehitetään käytänteitä mahdollistamaan datan jakamista ja hyödyntämistä**

Erilaisten kokeilujen tekemisen ja nopean sovelluskehityksen kannalta on ongelmallista, että iso osa niin sanotuista PaaS (*Platform as a service*) -palveluista, kuten Amazon Web Services, siirtää mahdollisesti henkilötietoja ETA-alueen ulkopuolelle. Näiden käyttäminen ei siksi ole tietosuojalainsäädännön vuoksi useinkaan mahdollisia. Tilastokeskuksen etäkäyttöpalvelua ja Valtorin palveluja kehitettäessä tulisi ottaa tämä huomioon ja pyrkiä edistämään kotimaisten tai EU:n sisäisten tiedon tallentamisen pilvipalvelujen kehittämistä niin, että PaaS-mallin pilvipalvelinympäristöjen edut saadaan datatalouden käyttöön ilman kansalaisten tietosuojan vaarantumista.

Yritysten ja kuluttajien (ns. "B2C") välisten käyttöehtojen ja mallisopimusten ohella tulisi kiinnittää huomiota siihen, miten yritysten välisessä toiminnassa ("B2B") saataisiin luotua malleja, joilla relevanttia dataa voidaan kerätä, anonymisoida ja jakaa tavalla, joka ei vaarantaisi yritysten liikesalaisuuksia. Toimivia käytänteitä yritysten välisten transaktioiden synnyttämän datan keräämisestä tulisi hakea ja jakaa aktiivisesti yritysten ja julkisen sektorin yhteistyönä.

## **Linjaus 2.3 Vaikutetaan valtion esimerkillä datan jakamisen synergiahyötyihin**

Aktiivinen, asioihin puuttuva analytiikka tekee data-aineistoista vähemmän geneerisiä ja siten kolmansille osapuolille vähemmän hyödyllisiä, ellei analytiikkaa nimenomaan suunnitella geneeriseksi. Tämän vuoksi datan jakamisessa tärkeää on kokonaisvaltaisuus, johon kuuluu niin datan keräämisen suunnittelu yhdessä sidosryhmien kanssa kuin sen pohjalta tehty edistynyt analytiikka.

Datan jakamisen ohella valtion tulisi selvittää mahdollisuutta julkistaa esimerkiksi koneoppimisella tuotettuja prediktiviisiä rajapintoja (API:t) Ilmatieteen laitoksen paikallissään tapaan (joka tosin perustuu enemmän fysikaaliseen malliin probabilistisen sijaan).

## **Linjaus 3.1 Datan soveltaminen yhteiskehittelyn ja yhteistyön kautta**

Pelkkä datan jakaminen tai kehitysympäristöjen ylläpitäminen ei takaa lisäarvoa tuottavien uusien datan hyödyntämiseen perustuvien palveluiden syntymistä. Näiden lisäksi vaaditaan datan jakajan toiminnan sekä dataa tuottavien ja hyödyntävien prosessien ymmärtämistä eli laajempaa ymmärrystä siitä liikatoimintakontekstista jossa data syntyy. Siksi esimerkiksi hackathon-tyyppisten tilaisuuksien lisäksi tarvitaan tiiviimpää yhteistyötä datan omistavan ja siitä palveluja kehittävän tahon välillä.

Erilaisissa tietoaineistoja hyödyntävissä hankkeissa alan suurimmat vaikeudet ovat ennen kaikkea tavoitteiden kommunikoinnissa liiketoimintojen sisällä ja varmaankin myös viranomaisten välillä sekä hyödyntäjien kesken. Olipa kysymys siis koulutuksesta, liiketoiminnan tukemisesta tai viranomaistoiminnasta, on keskityttävä analytiikkakulttuurin yhteiseen

jakamiseen, jos datavarannoille tosiaan halutaan MIT:n (Brynjolfson et al.) havaitsema 5-6% tuottavuusnousu.

Toiminnan esteet ovat yleensä muualla kuin itse datan puuttumisessa tai datatieteilijöiden menetelmäosaamisessa. Toiminnan esteet ovat pääosin systeemisiä: ratkaisuihin siirtyä pääpaino siirtyy datasta ratkaisuprosessiin, empiiriseen analytiikkakulttuuriin, empiiris-tilastollisen ajattelutavan tuomiseen kaikille tasoille arvoketjua. Empiiris-tilastolliseen ajattelutapaan perustuvien kehityshankkeiden ja toiminta pitäisi olla erityisen ketterää ja erilaiset organisaatiosiiilot ylittävää samoin kuin ohjelmistokehityksen.

### **Linjaus 3.2 Tehostetaan resursointia ja rahoituksen saatavuutta kansainvälisille markkinoille pääsyyn**

Alan ekosysteemeihin kohdentuvaa tutkimus- ja kehitysrahoitusta kehitettäessä tulisi muuttaa muun muassa Tekesin rahoitusmalleja niin, että ne mahdollistaisivat nykyistä ketterämmän rahoituksen ja matalamman byrokratian esimerkiksi pienille kasvuyrityksille. Tämä kannustaisi kasvuyritysten verkostoitumista ja siten ekosysteemien syntymistä.

Yritysten ja tutkimuslaitosten yhteistyötä pitää tehostaa niin, että opiskelijoiden ja jatko-opiskelijoiden on mahdollista saada ennen valmistumistaan jo käsitys työelämän vaatimuksista ja siellä tarvittavasta osaamisesta. Samalla uusia tutkimustuloksia saataisiin nopeasti yritysten hyödynnettäväksi.