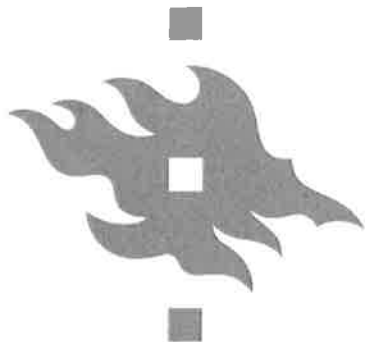




18.6.2014

Viestintä- ja liikenneministeriölle

Helsingin yliopiston lausunto kansallisesta big data -strategiasta

Helsingin yliopisto lausuu seuraavasti luonnoksesta kansalliseksi big data strategiaksi (LVM138/05/2014):

Helsingin yliopisto kannattaa kansallisen big data-strategian hyväksymistä, ja pitää lausuttavana olevaa luonnosta hyvänä lähtökohtana strategialle. Helsingin yliopisto katsoo kuitenkin, että tieteelliseen tutkimukseen liittyvät näkökulmat on nostettava strategiassa selkeämmin esiin.

Tutkimuksen merkitys big data-strategialle

Kansallisessa big data-strategiassa on tuotava ilmi, että ainoa yleisesti hyväksyttävä kriteeri tuotettavan tiedon laadulle on se, että tieto on tuotettu tieteellistä menetelmää käyttäen. Strategiassa todetaan, että big data -menetelmät ovat suurilta osin lähtöisin kaupalliselta puolelta (s.12). Tämä voi pitää paikkaansa joidenkin suosittujen analyysityökalujen osalta, mutta alan tutkimuksessa yliopistoilla on edelleen keskeinen rooli: monet big data-tutkimusmenetelmät ovat syntyneet tieteellisen tutkimuksen tarpeisiin. Useat niin luonnontieteissä kuin yhteiskuntatieteissä käytettävät tutkimusmenetelmät ovat olleet big datan tutkimusmenetelmiä jo ennen 'big data'-termin keksimistä. Helsingin yliopisto kyseenalaistaa strategialuonnoksessa esitetyn väitteen, jonka mukaan "big data-menetelmien käyttö on suomalaisessa tutkimuksessa vielä jokseenkin vähäistä" (s. 12). Iso osa Helsingin yliopistossa tehtävästä empiirisestä tutkimuksesta voitaisiin haluttaessa sijoittaa big datan sateenvarjon alle.

Perustutkimuksen näkökulmasta big data ei ole niin uusi asia kuin strategialuonnoksessa annetaan ymmärtää, eikä big data-aikakausi ole muuttanut tutkimusparadigmaa merkittävästi. Massadatan käyttö on lähinnä mahdollistanut tarkemmat mittaukset tai harvinaisempien ilmiöiden havaitsemisen.

Strategiassa ei huomioida riittävästi big datan erilaisia syntyprosesseja ja käyttötapoja tutkimuksessa. Eräillä tutkimusaloilla tuotetaan mittavia aineistoja. Esimerkiksi lääketieteellinen ja biotieteellinen tutkimus on muuttumassa voimakkaasti big data-tieteiksi, ja nämä alat kuuluvatkin johtaviin massadatan tuottajiin. Näillä aloilla syntyvän datan tallennus sekä datan tilastollisten ja laskennallisten analysointimahdollisuuksien turvaaminen on kansallinen haaste. Toisilla tutkimusaloilla, esimerkiksi yhteiskuntatieteissä, big data on puolestaan usein sekundaarista dataa eli se on kerätty ja tuotettu johonkin toiseen tarkoitukseen kuin mihin sitä halutaan käyttää. Tätä asiaa ei ole tuotu strategiassa esille, vaikka juuri se erottaa big datan analysoinnin tavallisesta empiirisen tutkimuksen prosessista. Ei voida itse rätätellä tapaa, jolla aineistoksi saataisiin havaintoja, jotka kuvastavat kiinnostuksen kohteena olevaa ilmiötä mahdollisimman tarkoituksemukaisella tavalla, vaan joudutaan tyytymään siihen dataan, mitä on saatavilla. Tällöin big datan avulla saadun tiedon hyödyntäminen vaatii tilastollisen päättelyn lisäksi esimerkiksi kontekstuaalista, kulttuurista ja historiallista tulkintaa.

Helsingin yliopisto oudoksuu strategialuonnoksessa esitettyä käsitystä siitä, että big data poistaisi tutkimuksesta tarpeen muodostaa teoriaa (s. 3). Kuten luonnoksessakin todetaan, big dataan liittyvien uhkien todennäköisyys kasvaa, jos siitä johdetut "päätelemät on tehty virheellisen tiedon tai mallien perusteella" (s. 4). Teorioilla ja malleilla on selvä rooli big dataa hyödyntävässä tutkimuksessa. Myös monissa käytännön sovelluksissa datamäärät eivät lopulta ole ongelma, vaan datan monimutkaisen rakenteen ymmärtäminen ja useiden epäyhtenäisten datalähteiden yhdistäminen.

Strategialuonnoksessa tulee kiitettävällä tavalla esille se, että big datan tulkinta ja käyttö edellyttää monitieteistä osaamista. Strategialuonnos ei kuitenkaan tunnista riittävästi eri tieteenalojen – humanistista tieteistä luonnontieteisiin – merkitystä big datan kehittymiselle ja hyödyntämiselle.

Big data-osaaminen, koulutus ja palvelut

Helsingin yliopisto toteaa, että luonnoksessa esitetty strateginen linjaus (big) dataan liittyvän koulutuksen lisäämisestä läpi koko koulutusjärjestelmän on tarkoituksenmukainen. Sen lisäksi, että tarvitaan tieteidenvälistä yhteistyötä ja tutkimusryhmiä, joissa on osaajia eri aloilta, jokaisen tutkijan on myös omaksuttava sovellettua informatiikkaa. Pelkkä big dataan liittyvä menetelmäkoulutus ei siis riitä. Menetelmä on aina vain työkalu, ei itseisarvo, ja tuloksia saa vain aineistonsa ymmärtävä tutkija. Tarvitaan henkilöitä, joilla on sekä substanssiosaamista että ymmärrystä teknisistä ratkaisumalleista, datanhallinnasta ja tilastotieteestä.

Samalla on kuitenkin huomattava, että silloinkin, kun tutkija on massadatan asiantuntija, hänellä on harvoin suurta intressiä tietovarannon koko elinkaareen. Sekä tutkimuslaitoksissa että julkishallinnossa pitäisikin kehittää datan hallinnoinnin ratkaisuja, jotka eivät liikaa rasita yksittäisiä tutkijoita ja tutkimusryhmiä. Strategialuonnos käsittelee infrastruktuureita lähinnä teknologioina. Olisi kuitenkin asianmukaista tuoda esiin, millaisia palveluita ja osaamisresurssia voisi kuulua massadatan hyödyntämistä tukevaan infrastruktuuriin. Helsingin yliopisto esittää, että strategiassa kiinnitetään huomiota big dataan liittyvien palvelun tarjoajien, palveluiden käyttäjien ja palveluiden rahoittajien välisen rakentavan kolmikantaisen yhteistyön ja vastuunjaon luomiseen.

Tutkimusrahoituksen kohdentaminen

Helsingin yliopisto katsoo, että tutkimusrahoitusta on ohjattava yleisemmällä tasolla perustutkimukseen, jossa käytetään hyväksi big dataa ja siihen liittyviä menetelmiä. Strategialuonnoksen sanamuoto (s.14) antaa ymmärtää, että rahoitusta ohjaamalla voidaan suosia teknologiaosaajien ja yhteiskuntatieteilijöiden yhteistyötä big dataan liittyvissä asioissa. Tässä tulisi huomioida selkeästi myös muut tieteenalat.

Helsingin yliopisto kannattaa ehdotusta, että Tekes kohdistaisi ohjelmarahoitustaan big data-alueelle. Lisäksi yliopisto esittää, että myös Suomen Akatemia suuntaa rahoitusta tälle alueelle, esimerkiksi ICT-2023 -ohjelmansa ja uuden Strategisen tutkimuksen neuvoston kautta.

Big datan hyväksikäytön juridiset haasteet

Helsingin yliopiston mukaan kansallisen big data-strategian tulee edistää ei-kaupallista tutkimuskäyttöä koskevan poikkeussäännöksen sisällyttämistä kansalliseen tekijänoikeuslakiin.

Tämä kehittäisi huomattavasti digitaalisten aineistojen laajempaa hyödyntämistä tutkimuksessa ja tehostaisi tutkimustyötä.

Strategialuonnoksessa big data käsittää pääasiassa sellaisen datan, johon ei liity tekijänoikeuskysymyksiä. Näkökulmaa on laajennettava. Tekijänoikeuslaki on muodostunut muun muassa kieliaineistojen käytön ja tiedonlouhinnan esteeksi. Liian ankarat tekijänoikeuskäsitteet hidastavat tai suorastaan estävät tutkimusta, joka olisi välttämätöntä yhteiskunnan kehitykselle ja yhteiseksi eduksi. Luonnoksessa todetaan, että Euroopan ja Suomen big data-kehitys on erityisesti USA:ta jäljessä. Tämä johtuu osittain siitä, että tutkimusta ei ole mietitty riittävästi EU-lainsäädännössä.

Poikkihallinnollisuus ja kansallinen yhteistyö

Helsingin yliopisto korostaa, että kansalliseessa big data -strategiassa on huomioitava selkeämmin yhteys opetus- ja kulttuuriministeriön Avoin tiede ja tutkimus (ATT) -hankkeeseen. Strategialuonnoksessa korostetaan poikkihallinnollisuutta (s. 7). Käytännössä tämä tavoite toteutuu kuitenkin huonosti, muun muassa sen takia, että ATT-hanketta ei ole otettu huomioon. ATT-hankkeessa etsitään ratkaisuja big data-strategialuonnoksessa esitettyihin ongelmiin (datan tallennus, pitkäaikaissäilytys ja jakaminen, standardointiratkaisut, metatietomalli, datan jakamista koskevat sopimukset). Suomen kokoisessa maassa ei ole esimerkiksi järkevää rakentaa useita tutkimusdatan pitkäaikaissäilytysjärjestelmiä eikä kehittää useita rinnakkaisia standardointiratkaisuja pitkäaikaissäilytystä varten.

Helsingin yliopistossa toteutettavat big dataan liittyvät toimenpiteet ja Helsingin yliopiston rooli big data-kehittämistyössä

Helsingin yliopistolla on erinomaiset edellytykset toimia aktiivisesti big data-osaamisen ja tutkimuksen keskittymänä sekä tarkoituksenmukaisen koulutuksen tuottajana. Alla on esitetty muutamia esimerkkejä tästä.

Helsingin yliopiston humanistinen tiedekunta on edistänyt massadataan liittyvien uusien mahdollisuuksien käyttöönottoa perustamalla digitaalisten aineistojen professuurin, joka kattaa kielellisten ja muiden humanististen aineistojen käytön tutkimuksen sekä humanistisen alan tutkimusaineistojen uudenlaisen käytön ja uudenlaisen tiedon luomisen. Professori Timo Honkela aloitti tehtävässään vuoden 2014 alussa.

Nykykielten laitoksella on vahvaa laajojen tekstiaineistojen tutkimusosaamista ja opetusta korpuslingvistiikassa ja kieliteknologiassa. Laitoksella on myös vastuu kieliaineistojen ja kieliteknologian yhteisestä tutkimusinfrastruktuurien tiekarttahankkeesta, FIN-CLARIN-konsortioista.

Aleksanteri-instituutti toimii suunnannäyttäjänä Venäjän julkisen sektorin tuottaman avoimen datan sekä verkkoteknologioiden automaattisesti generoiman big datan tutkimuksellisessa hyödyntämisessä. Tietojen jäsentelemisen kautta voidaan tarjota kansainväliselle Venäjän- ja Itä-Euroopan tutkijayhteisölle kohdennettua palvelua.

Valtiotieteellisessä tiedekunnassa on vahvaa osaamista rekisteriaineistojen analysoinnista, ja siellä tehdään aiheeseen liittyvää menetelmällistä tutkimusta (erityisesti rekisteriaineistojen osalta) yhteistyössä muun muassa Tilastokeskuksen, THL:n ja Kelan sekä muiden maiden tilastovirastojen kanssa. Big data-teema on huomioitu tiedekunnan tohtorikoulutuksessa.

Lääketieteellisessä tiedekunnassa ja Suomen molekyyli lääketieteen instituutissa (FIMM) työskentelevät tutkijat johtavat terveyteen liittyvää genomitutkimusta, joka on lähtökohtaisesti big data-tutkimusta. Tutkimuksessa syntyvän tiedon tallennus sekä tiedon tilastollisten ja laskennallisten analysointimahdollisuuksien turvaaminen tulee olemaan keskeistä kansallisen terveystutkimuksen näkökulmasta. Lääketieteellisellä tiedekunnalla ja FIMM:lla onkin asiantuntemusta osallistua kansalliseen strategiseen kehitystyöhön niin terveystutkimuksen kuin big data-strategioiden osalta (tallennus, tilastolliset ja laskennalliset analyysimahdollisuudet, erityisesti lääketieteen yhteiseurooppalaiset tutkimusinfrastruktuurit). Lääketieteellisessä tiedekunnassa ja FIMM:ssa on myös asiantuntemusta sensitiivisten aineistojen käyttöoikeuksien hallinnan ja lupaprosessien kehittämistyöhön.

Tietojenkäsittelytieteen laitos tuottaa koneoppimisen ja algoritmien tutkimusta. Ilman tätä tutkimusta big data menettää merkityksensä. Laitos voi kehittää big data-algoritmeja yhteistyössä dataa keräävien ja analyyseja tuottavien tahojen kanssa. Näin saadaan aikaan sovellusalan erityistarpeet huomioonottavia analyysimenetelmiä.

Helsingin yliopisto on Suomen suurimpia tietojenkäsittelytieteen alan kouluttajia, ja sellaisena halukas suuntaamaan koulutustaan alan tarpeiden mukaisesti. Big data-opetus kytkeytyy yleisempään tietojenkäsittelytieteen, tilastollisen data-analyysin, matemaattisen mallinnuksen ja hajautetun laskennan opetukseen, jota Helsingin yliopistossa on annettu jo kauan ennen kuin termi big data tuli yleiseen kielenkäyttöön. Tietojenkäsittelytieteen laitos käynnistää lukuvuonna 2014-2015 uuden maisteriopintojen erikoistumisprofiilin, jossa erikoistutaan datatieteeseen ja big dataan. Profiili yhdistää data-analyysin ja nykyaikaisen Internet-laskentaympäristön taidot.

Helsingin yliopisto on tehnyt pitkään yhteistyötä muiden yliopistojen kanssa ja tarjonnut tilastollisen data-analyysin osaamista tapauskohtaisesti kunkin alan tarpeeseen. Yliopistossa on esimerkiksi toiminut vuodesta 2002 alkaen useita valtakunnallisia Suomen Akatemian huippuyksikköjä, joiden voi katsoa kuuluvan big data-teemaan. Helsingin yliopiston ja Aalto-yliopiston big data-osaaminen yhdistyy Tietotekniikan tutkimuslaitoksessa (HIIT).

Helsingin yliopisto valmistelee tutkimusdatapolitiikkaansa. Poliitiikan on määrä valmistua vuoden 2015 alussa. Poliitikassa sovitaan tutkimusdatan keräämistä, käsittelyä, tallentamista ja säilyttämistä koskevista yliopistotason linjauksista ja periaatteista. Poliitiikan valmistelussa huomioidaan kansalliset strategiat ja pyritään aktiivisesti vaikuttamaan kansallisiin kehityslinjoihin.

Helsingin yliopisto monialaisena yliopistona antaa erinomaiset puitteet tieteidenväliselle yhteistyölle, joka hyödyntää ja edelleen kartuttaa eri alojen asiantuntemusta laajojen aineistojen analyysissä.



Keijo Hämäläinen
Vararehtori



Eeva Nyrövaara
Tutkimushallinnon asiantuntija