

minedu.fi/DL2021

Datanhallinnan ja laskennan kehittämishjelman ohjausryhmä

Aika: torstaina 11.1.2018 klo 13–15

Paikka: opetus- ja kulttuuriministeriö, kh. Kalevala, Meritullinkatu 10

Ohjausryhmän kokoonpano, läsnäolo

tiedeasiantuntija *Erja Heikkinen*, opetus- ja kulttuuriministeriö, pj.
erityisasiantuntija *Teemupekka Virtanen*, sosiaali- ja terveysministeriö
kaupallinen neuvos *Antti Eskola*, työ- ja elinkeinoministeriö
viestintäneuvos *Anne Miettinen*, liikenne- ja viestintäministeriö
professori *Sampsa Hautaniemi*, Helsingin yliopisto
vararehtori *Kirsi Viskari*, Saimaan ammattikorkeakoulu
ohjelmapäällikkö *Sirpa Thessler*, Luonnovarakeskus
pääjohtaja *Heikki Mannila*, Suomen Akatemia
kehityspäällikkö ~~*Tiina Tanninen-Ahonen*~~, Business Finland, este
tutkimusjohtaja *Esa Virtanen*, ABB Oy
professori *Paula Eerola*, Helsingin yliopisto
opetusneuvos *Petteri Kauppinen*, opetus- ja kulttuuriministeriö

Asiantuntijajäsenet

Opetusneuvos *Juha Haataja*, opetus- ja kulttuuriministeriö
Ylitarkastaja *Sami Niinimäki*, opetus- ja kulttuuriministeriö

Johtava tiedeasiantuntija *Anu Nuutinen*, Suomen Akatemia (asiantuntijas sihteeri)

Kutsut vieraat

varatoimitusjohtaja *Tiina Kupila-Rantala*, CSC
rehtori/toimitusjohtaja *Turo Kilpeläinen*, KAMK; läsnä asiakohdissa 1–3

1. Kokouksen avaus ja asialistan hyväksyminen

Puheenjohtaja avasi kokouksen ja etukäteen toimitettu **asialista hyväksyttiin**.

TEM ilmoitti vaihtavansa varajäsentä (johtava asiantuntija *Jonna Lehtinen-Salo*) nykyisen varajäsenen siirtyessä toisiin tehtäviin. Sovittiin, että uuden varajäsenen yhteystiedot ilmoitetaan sihteeristölle.

2. Edellisen kokouksen muistio ja kehittämissohjelman valmistelu syksyllä 2017

Edellisen kokouksen muistio hyväksyttiin. Ohjausryhmälle oli toimitettu tiedoksi myös taustamuistio DL2021-kehittämissohjelman valmistelusta syksyllä 2017 sekä CSC:n laskentaresurssin käytöstä ja Suomen Akatemian rahoituksesta taustaselvitys, jossa vertailua tehtiin tieteenaloittain ja organisaatioittain.

minedu.fi/DL2021

3. Tutkimuslaitos selvitys

Kehittämisohjelman toteutuksen tueksi tehtiin selvitys, jossa kuvataan valtion 12 tutkimuslaitoksen datanhallinnan ja laskennan sekä tietoliikenneyhteyksien rooli/tarve ja tulevaisuuden tavoitetilä. Selvityksessä on kartoitettu myös laskentakapasiteetin yksilöityä tarvetta. Lopuksi esitetään toimenpide-ehdotuksia, jotka sisältävät myös esityksen kansallisiksi datanhallinnan ja laskennan piloteiksi.

Turo Kilpeläinen esitteli selvityksensä keskeiset tulokset ja toimenpide-ehdotukset. Kilpeläinen korosti kansallisen yhteistyön tärkeyttä. Osaaminen on keskeinen asia Suomen data- ja laskentaintensiivisen tieteen kokonaisratkaisussa. Dataintensiivisen tieteen tarpeet on keskeistä ottaa huomioon hankinnassa.

Tutkimuslaitosten osalta todettiin, että selvitys antaa osin aika karun mutta rehellisen kuvan datanhallinnan ja laskennan tilanteesta. Koska nykyisin käytetään paljon ostopalveluja tietovirtojen hallinnassa, osaaminen ei kumuloidu tutkimuslaitoksiin. Teollisuudessa puolestaan tilanne on toisenlainen, nykyisin ostetaan kaikki ulkoa; kansainvälinen näkökulma korostuu.

Keskustelussa korostettiin, että infrastruktuuri pitää saada ensin kuntoon ennen kuin voidaan keskustella tekoälyn hyödyntämisestä. Pullonkaulana on perusinfrastruktuurin kunto. Data-analyysi ja tekoäly ovat toistensa jatkumoa. Osaajia tarvitaan kaikilla tasoilla.

Kysyttiin myös, miten tulevaisuuden laskentatarpeet (vuoden 2021 jälkeen) on otettu huomioon. Turo Kilpeläinen totesi, että tulevaisuuden tarpeiden ennakoiminen osoittautui tutkimuslaitoksissa erittäin vaikeaksi: ei ole mahdollista nähdä kovin tarkasti 10 vuoden päähän, kun nykyisenkin data- ja laskentaintensiivisen tutkimuksen volyymin arvioimisessa on haasteita.

Puheenjohtaja totesi, että kehittämisohjelman avulla pyritään nimenomaan rakentamaan osaamista kansallisesti sen sijaan, että ostetaan kaikki palvelut ulkopuolelta. Kehittämisohjelmassa päästään hahmottelemaan myös seuraavan infrastruktuurisyklin valmistelua.

Ohjausryhmä kannatti lähestymistapana pilotointia, jossa tehdään konkreettista kehitystyötä. Tunnistettiin, että ehdotettujen pilottien lista on itse asiassa pitkälti samannäköinen kuin infrastruktuurien kansallinen tiekartta. Tiekartan tutkimusinfrastruktuureiden väliarviointi on parhaillaan käynnissä.

Keskusteltiin myös pilottien rahoituksesta. Puheenjohtaja painotti, että jokaiselta hallinnonalalta tarvitaan panostuksia. Valmistelussa on hallinnonalojen kansliapäällikköjen tapaaminen, jossa DL-asiakokonaisuutta käsitellään.

Sovittiin, että toteutetaan pilotteja ja että CSC ja tutkimuslaitokset sopivat keskenään niiden toteutuksesta käytännössä.

minedu.fi/DL2021

Puheenjohtaja myös totesi, että ohjausryhmällä oli mahdollisuus kommentoida sopimusluonnosta Datanhallinnan ja laskennan kehittämisohjelman toteuttamisen edellyttämän osaamisen ja tukipalveluiden rahoittamisesta vuosina 2018–2019 sähköpostitse. Joulukuussa 2017 allekirjoitetulla sopimuksella osoitettiin rahoitusta sekä CSC:n oman että laajenevan asiakaskunnan osaamisen kehittämiseen.

4. Datanhallinnan ja laskennan kehittämisohjelman hankintaesitys

Tiina Kupila-Rantala esitteli ohjausryhmälle Datanhallinnan ja laskennan kehittämisohjelman hankintaesityksen. Kehittämisohjelman avulla ratkaistavia asiakastarpeita voidaan jäsentää seuraavien kuuden käyttötapauksen kautta:

- Suuren skaalan simulaatiot
- Keskikokoiset simulaatiot
- Dataintensiivinen laskenta
- Dataintensiivinen laskenta sensitiivisellä datalla
- Tekoäly
- Internet of Things ja datavirrat.

Hankinnan vaiheistuksen periaatteet ovat seuraavat:

- Täytyy osata arvioida, kuinka pitkään kannattaa hyödyntää olemassa olevaa infrastruktuuria eli optimoida olemassa olevien laitteiden käyttöikä.
- Rahoitusta käytetään ensisijaisesti laskentatehoon ja datanhallinnan instrumentteihin, mahdollisimman vähän laitetilaan.
- Suunnittelussa pyritään huolehtimaan, että ei tule katvealueita palvelutarjontaan millekään alalle, kun fyysiset koneet vaihtavat paikkaa.

Infrastruktuuri asennetaan kahdessa päävaiheessa. Ensimmäisessä vaiheessa laskentakapasiteetti tulee asiakkaiden käyttöön maaliskuuhun 2019 mennessä, toinen vaihe valmistuu maaliskuussa 2020. Yksi palvelutarjonnan kriittinen vaihe asiakaskunnan odotuksien hallinnassa tulee olemaan vuoden 2019 loppupuolella, kun Sisun paikalle asennetaan uusi kone.

Osaamisen kehittäminen käynnistyy 2018. Uusien käyttäjäryhmien (tutkimuslaitokset ja tieteenalat) kanssa käydään läpi mm. verkkokurssitarpeita, data-aineistojen hallintaa, menetelmäkysymyksiä sekä esim. konttiteknologioihin liittyviä kysymyksiä (konttiteknologia viittaa paketoititekniikkaan, jolloin sovellusta ei tarvitse räätälöidä tiettyyn ympäristöön). Sensitiivisen datan käsittely järjestetään muulla rahoituksella pilvipalveluiden kanssa.

Ohjausryhmä keskusteli hankintaehdotuksesta. Keskusteltiin dataintensiivisen laskennan ja erityisesti sensitiivisen datan huomioimisesta uuden infrastruktuurin hankinnassa ja hankinnan vaiheistuksessa sekä siitä, mitkä tulevat olemaan keskeisimmät muutokset uudessa ympäristössä

minedu.fi/DL2021

nykytilanteeseen verrattuna. Muistutettiin, että lähivuosina lainsäädäntö tulee vaikuttamaan ihmisperäisten aineistojen laskennalliseen tutkimukseen merkittävästi.

Todettiin, että kaupallisessa pilvessä on vaikea toteuttaa 10 000 ja yli laskentaydintä hyödyntävää laskentaa.

Asiakastarpeet voidaan nähdä jatkumona. Data-access on ollut pullonkaula. Hankintaesityksessä on nyt otettu huomioon nopea työlevy, joka on keskeinen suurissa ja keskisuurissa simulaatioissa.

Investointisuunnitelmaan toivottiin kuvausta myös siitä, miten asiantuntijuuden kehittäminen näkyy suunnitelmassa. Puheenjohtaja viittasi asiakohdassa 3 kuvattuun päätökseen osaamisen kehittämisestä, johon voidaan viitata tätä hankintaa koskevassa sopimuksessa.

Ohjausryhmä kysyi, kattaako datanhallinta esim. mobiilidatan. Sopimus tulee kattamaan kaikenlaisen datanhallinnan. Onko mahdollista ottaa yleislaskentasuorituksiin myös sensitiivinen data huomioon? Tätä ei ole kategorisesti poissuljettu. Tärkeää on kuitenkin korostaa, että datan turvallisuus syntyy myös kehittyvistä toimintatavoista.

Tiina Kupila-Rantala korosti, että hankittavan infrastruktuurin täytyy pystyä palvelemaan monenlaisia tarpeita. Keskeinen muutos nykytilanteeseen on se, että hankintaesityksessä pyritään saamaan infrastruktuuri, joka tukee myös dataintensiivisiä tarpeita. Aiemmin yhteinen tallennustila on palvellut lähinnä suurinta laskentaa.

Puheenjohtaja tähdensi, että OKM investoi myös muuten kuin DL2021-rahoituksen kautta CSC:n DL-palveluiden kehitykseen. **Kilpailutukseen kuulumattomat palvelut ja niihin liittyvät osat uudistetaan muulla rahoituksella (sensiitivinen data).**

Ohjausryhmä antoi OKM:lle valtuutuksen sopia CSC:n kanssa DL2021-hankinnasta osana vuosisopimusta. Ohjausryhmän käsittelemästä Datanhallinnan ja laskennan kehittämisohjelman hankintaesityksen tiivistelmästä tulee sopimuksen liite.

5. Eurooppalainen tutkimusinfrastruktuuripolitiikka datanhallinnan ja laskennan infrastruktuurien näkökulmasta

Asiakohta siirrettiin käsiteltäväksi seuraavassa kokouksessa.

6. Seuraavat kokoukset

STM ehdotti, että hyvinvoinnin palveluoperaattori Isaacusta käsitellään sopivassa ohjausryhmän kokouksessa, kun sote-tiedon toisiokäyttöä koskeva laki on valmistunut.

Ohjausryhmän 3. kokous 16.4.2018 klo 9–11 (kokousajat on vahvistettu, aiheista on alustava, muutokset ovat mahdollisia)

- CSC:n DL-hankintaa koskevan kilpailutuksen eteneminen

minedu.fi/DL2021

- Koulutuksen ja tutkimuksen aktivointi datanhallintaan ja laskentaan liittyen / Osaamisen kehittäminen
- Miten tutkimuslaitosten tilanne CSC:n asiakkaana on kehittynyt?
- Tutkimustietovarannon valmistelun eteneminen
- TEAS-tekoälyhankkeen esittely (hankkeen toteuttajaa ei ole vielä julkistettu)
- Yksityisen sektorin tarpeet – Kartoituksen tuloksia

Ohjausryhmän 4. kokous 27.9.2018 klo 13–15

- CSC:n DL-hankintaa koskevan kilpailutuksen eteneminen
- Ministeriöiden tilannekuvat toimialaltaan (yhteistyö tutkimustoimijoiden ohjauksessa ja eri tahojen seuranta kehittämisohjelman tavoitteiden saavuttamisessa)
- Datanhallintaan vaikuttavan lainsäädännön uudistusten toimeenpanon eteneminen