

minedu.fi/DL2021

Datanhallinnan ja laskennan kehittämishjelman ohjausryhmä

Aika: tiistaina 17.9.2019 klo 10.00–12.00

Paikka: OKM, Meritullinkatu 10, Kh. Louhi

Ohjausryhmän kokoonpano ja varajäsenet

johtaja *Erja Heikkinen*, opetus- ja kulttuuriministeriö, pj.
erityisasiantuntija *Teemupekka Virtanen*, sosiaali- ja terveysministeriö
neuvotteleva virkamies *Pirjo Kutinlahti*, työ- ja elinkeinoministeriö (asiantuntija *Maikki Sipinen*)
erityisasiantuntija *Lotta Engdahl*, liikenne- ja viestintäministeriö (neuvotteleva virkamies *Tuomas Kaivola*)
professori *Sampsa Hautaniemi*, Helsingin yliopisto
TKI-vararehtori *Kirsi Viskari (etänä)*, Tampereen ammattikorkeakoulu
ohjelmapäällikkö *Sirpa Thessler*, Luonnonvarakeskus
pääjohtaja *Heikki Mannila*, Suomen Akatemia
Head of AI Business *Outi Keski-Äijö*, Business Finland Oy
tutkimusjohtaja *Esa Virtanen*, ABB Oy
professori *Paula Eerola*, Helsingin yliopisto (professori *Petri Myllymäki*)
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)

Vierailijat

varatoimitusjohtaja *Tiina Kupila-Rantala*, CSC
johtaja *Pekka Lehtovuori*, CSC

minedu.fi/DL2021

1. Kokouksen avaus ja asialistan hyväksyminen

Puheenjohtaja avasi kokouksen ja asialista hyväksyttiin.

2. Edellisen kokouksen muistio

Kokousmuistio 15.5.2019 pidetystä ohjausryhmän kokouksesta hyväksyttiin.

3. Yritysten laskentatarpeet

Outi Keski-Äijö Business Finlandista esitteli kyselyä, jossa oli selvitetty yritysten odotuksia CSC:n datalaskennalle ja muille palveluille. Lisäksi esitystä varten oli haastateltu paria laskentaa runsaasti hyödyntävää yritystä.

Yritykset tarvitsevat yksilöllisiä ratkaisuja. SaaS-mallilla toimivat yritykset ovat määrällisesti suuri joukko, jonka palveleminen olisi vain suuri rasite CSC:lle yrityksille toteutuneeseen arvoon nähden. AI:ta soveltavat yritykset, joilla on tuotekehitysvaiheessa korkea laskentatarve, kokevat kaupalliset pilvipalvelujen tarjoajat houkuttelevimmiksi. Vaikka kustannukset ovat isot, saadut alennukset vähentävät toteutuneita kustannuksia, työkalut ovat hyviä, palvelu pyörii 24/7 ja saatavilla on palveluja pilven päälle, mikä voi poistaa oman osaamisen ja resurssien tarvetta. Laskentaa runsaasti hyödyntävistä yrityksistä osa voisi olla CSC:n kohderyhmää erityisesti aikaisessa tutkimusvaiheessa, mutta näiden yritysten mukaan CSC:n ei kannata lähteä kilpailemaan kaupallisten tarjoajien kanssa samoilla palveluilla. Kaupallisten toimijoiden yksityiset pilvet koetaan myös tietoturvalisiksi, joten datan arkaluonteisuuskaan ei näytä, ainakaan laajasti, vaikuttavan päätöksiin kaupallisen toimijan käytöstä.

Ohjausryhmä kiitti selkeästä alustuksesta. Tuloksia ei pidetty yllättävinä. Keskusteltiin siitä, että suomenkieliset datasetit ovat tärkeitä (tekoälyä hyödyntäville) suomenkielisille yrityksille. Yrityksiä kiinnostaisi eniten sellaiset datasetit, joita ei saa muualta, ja että laskennan voisi tehdä datan kanssa samassa ympäristössä. Todettiin, että tällä hetkellä on puute yhteistyöalustoista, jossa esim. tutkijan algoritmi ja yrityksen data voitaisiin joustavasti yhdistää. Pullonkaula yrityksille näyttäisi olevan datan ymmärtäminen, ei niinkään laskentakapasiteetti, algoritmit tai datan säilytys. Rahoitusta pitäisi suunnata aitoihin yritys-tutkija-hakuihin.

CSC korosti, että datan käsittely on nimenomaan osaamiskysymys ja DL2021-kehittämishjelman kompetenssirahoituksella on rakennettu mm. Datasets as a Service -palvelua.

Outi Keski-Äijö totesi lopuksi, että kehitystyössä pitää edetä kokeillen, eikä palveluja voi speksata kerralla valmiiksi. Palveluiden räätälöinti on keskeistä. Puheenjohtaja totesi, että tähän aihepiiriin palataan vielä uudestaan.

4. DL2021-käyttöönoton ja osaamisen kehittämisen ajankohtaiset asiat CSC:llä

Pekka Lehtovuori CSC:ltä kävi läpi DL2021-tilannekatsauksen sekä esitteli uusimpia tilastoja CSC:n resurssienkäytön jakautumisesta käyttäjien ja tieteenalojen suhteen vuonna 2019. GPU-resurssit ovat kansainvälisesti erittäin kilpailukykyisiä (kiitos lisärahoituksen), ja ne rakentavat siltaa myös EuroHPC:n hyödyntämiseen.

minedu.fi/DL2021

Kevään 2019 aikana järjestettiin 15 tapahtumaa ja osallistuttiin 4 asiakasseminaariin DL2021-agendalla. Lisäksi järjestettiin yksi webinaari. Tapaamisissa oli kaksi eri agenda: toinen räätälöity uusille ja toinen vanhoille asiakkaille. Syksyllä 2019 keskitytään erityisesti ammattikorkeakouluihin. Keväälle 2020 suunnitellaan kampusseminaaria (yliopisto, ammattikorkeakoulut ja tutkimuslaitokset yhdessä) eli tavoitteena on edelleen vahvistaa eri toimijoiden yhteistyötä.

UNIFIn, ARENEn ja TULANETin kanssa suunnitellaan näiden järjestöjen omien verkostojen hyödyntämistä siten, että DL2021- ja EuroHPC-ohjelmien mahdollisuuksia voidaan hyödyntää täysimääräisesti ja maksimoida palveluiden tunnettuutta. Valtorin kanssa puolestaan on edetty hyvin teknisten käyttöesteiden poistamiseksi ja Valtorin tarjoamien valmiiksi paketoitujen työpöytäsovellusten saamiseksi tutkimuslaitosasiakkaiden koneisiin. Tutkimuslaitospiloteista esiteltiin Automatisoidut suuren skaalan datavirratt, Datasets as a Service -konsepti, Sensitiivisten aineistojen etätyöpöytä ja Automatisoidut IoT-datavirratt. Tutkimuslaitospiloteissa haasteita on ollut odotusten hallinnassa, kun huomattiin, että rahoitusta olisi tarvittu myös tutkimuslaitospuolelta. CSC ei voi toteuttaa kaikkia mahdollisia toiveita yksin.

Ohjausryhmä piti osaamisen kehittämisen pilottihankkeita mielenkiintoisina. Pohdittiin, sopsisiko joku pilotointi paremmin THL:lle kuin CSC:lle; miten esimerkiksi järjestetään se, että THL:n altaan tietoja voisi käyttää CSC:n ympäristössä.

Data as a service -palvelun tarkoitus on vastata myös yritys-korkeakouluyhteistyöhön. Todettiin, että yritykset eivät välttämättä laittaisi kaikkea dataansa sellaisenaan saataville, mutta dataa pitää olla riittävästi, jotta analyysit tuottavat realistisia tuloksia. Tätä pitää pilotoida. Kehitystyö liittyy aikaisen vaiheen toimintaan, kun riskit ovat isot eikä vielä ole varmuutta, tuleeko tuotteesta koskaan kaupallista toimintaa. Puheenjohtaja myös selvensi CSC:n omistajaohjauksen näkökulmaa: hankintadirektiivin kansallinen tulkinta rajaa laskentaresurssien yksityistä käyttöä. Tapoja, joilla voitaisiin ratkaista ja laajentaa kaupallista käyttöä, on identifioitu.

Ohjausryhmä kysyi amk-kiertueesta. AMK-tutkimusjohdon verkosto olisi hyvä paikka jakaa tietoa. Esittelyä on pyydetty verkoston kokoukseen marraskuun lopulla. TKI-valiokunnan kautta löytyy kiinnostuneita paikkakuntia. Koulutuspaketti on hyvä johdantopaketti. Onko sitä markkinoitu laajemmin muillekin kuin Kajaanin ammattikorkeakouluille?

5. EuroHPC, kompetenssikeskushaku ja HPC-ekosysteemin rakentaminen

Pekka Lehtovuori esitteli EuroHPC:n ja LUMI:n merkitystä tiedeyhteisölle ja sidosryhmille. Suuri lisäarvo tulee LUMI-konsortioista (9 maan yhteistyö) ja Kajaanin datakeskuskonseptista (energiansäästönäkökulmat ja kustannustehokkuus). Komissio haluaa lähentää myös EuroHPC:tä ja EOSCia. Lumi tulee olemaan yksi maailman näkyvimpiä tieteellisiä instrumentteja elinkaarensa 2020–2026 ajan. Konsortio luo käytänteet ja yhteistyön perinteen tuleville EuroHPC-laitteistosukupolville. LUMI tarjoaa maailmanluokan resurssit suomalaisille tutkijoille, mikä luo merkittävän kilpailukykyedun. LUMIn resurssit tulevat olemaan myös merkittävä vetovoimatekijä ulkomaisia tutkijoita rekrytoitaessa.

DL:n pitää pystyä palvelemaan kaikkien tieteenalojen laskennallisia tarpeita. EuroHPC:llä pitää lisäksi pystyä vastaamaan Horisontti Eurooppa -ohjelman missioihin. Henkilötyötä pitää satsata siihen, että kapasiteetti on hyödynnettävissä yliopistojen strategisten valintojen tueksi. LUMIn kautta voidaan myös osallistua isoihin kansainvälisiin infrastruktuureihin laskennan ja datanhallinnan osalta. Tutkimuksellisten tavoitteiden lisäksi

minedu.fi/DL2021

EuroHPC:n yhteiskunnallinen vaikuttavuus pitää pystyä osoittamaan. Elinkeinoelämän ja tutkimuksen yhteiskäyttö on tärkeää.

Ohjausryhmä korosti, että on tärkeää puhua EuroHPC:n rekrytointivaikutuksista. EuroHPC on tärkeä tekijä siinä, että eurooppalaiset tutkijahuiput pidetään Euroopassa.

EuroHPC:n laskentakapasiteetti yhdistetään myös DL-altaaseen, jolloin kansalliset tietovarannot saadaan käyttöön.

Keskusteltiin myös siitä, toteutetaanko kaikki kolme EuroHPC-konetta samoilla spekseillä. Pekka Lehtovuori tarkensi, että jo hakemusvaiheessa keskukset ovat profiloituneet. Komissio on velvoittanut, että keskukset eivät ole kilpailevia keskenään, vaan tarjoavat kattavasti Euroopassa tarvittavat huipputason HPC-resurssit.

EuroHPC-haut (tätä kokouksessa käsiteltyä osuutta on täydennetty muistioon kokouksen jälkeen myös muista lähteistä.)

- EuroHPC pre eksa -supertietokoneiden hankintamenettely toteutettiin 28.11.2019–3.1.2020.
- Viisi TKI-hankehakua teknologian kehittämiseksi avautui 25.7.2019. Hauista kaksi (teemasta Innovating and Widening the HPC use and skills base) sulkeutui 14.11.2019 ja kolme (teemasta Towards Extreme Scale Technologies and Applications) 14.1.2020.
- Marraskuussa sulkeutuneista hauista toinen keskittyy kompetenssikeskusten rahoittamiseen. Kompetenssikeskusten rahoitus toteutetaan kahtena hankkeena. Toisessa on mukana kunkin maan kompetenssikeskus omana työpaketinään. Toisessa hankkeessa on kokonaisuuden koordinointi. CSC on Suomen kansallinen kompetenssikeskus sekä LUMI-konsortion kompetenssikeskus.

EuroHPC-laskentakapasiteetin jakoperusteet

- EuroHPC-resurssien jaosta ei ole vielä päätetty, ovatko EuroHPC-laskentaresurssit käytössä samoilla periaatteilla kuin kansalliset DL-resurssit. Kokousajankohtana ei ollut tietoa, tuleeko JU:sta mahdollisesti erityisohjeistusta tähän.
- (Täydennetty kokouksen jälkeen: PRACE:lla on tärkeä rooli erityisesti uusien supertietokoneiden käyttöajan hallinnoinnissa (mm. vertaisarvioinnin järjestäminen). PRACE on palvellut erityisesti tiedeyhteisöä ja EuroHPC:ssa tulee huomioida myös teollisuuden osallistuminen. Nämä menettelyt olisi hyvä saada sovittua EuroHPC-JU:n ja PRACEn kesken kesäkuuhun 2020 mennessä.)
- Puheenjohtaja totesi, että lähtökohtana on, että CSC päättää isoja asiakkaita ja omistajia kuullen, miten kansallinen EuroHPC-laskentakapasiteetti jaetaan. Ei ole tarpeellista perustaa tätä varten uusia ohjausrakenteita. DL2021-ohjausryhmä toimii myös EuroHPC:n ohjaajana. DL-ohjausryhmässä ovat edustettuina OKM ja TEM eli EuroHPC:tä Suomesta rahoittavat tahot.

Puheenjohtaja kannusti ohjausryhmää hyödyntämään kaikkia mahdollisia yhteistyöverkostojaan aktiivisesti EuroHPC:n mahdollisuuksien markkinoinnissa. Faktat tulee olla markkinoinnissa kohdallaan. Aiheeseen tullaan palaamaan myös DL-ohjausryhmän kokouksissa.

6. DL2021-vuosiseminaarin valmistelu: Ajankohta, paikka, teema ja kohderyhmä

Anu Nuutinen esitteli vuosiseminaarin teeman, keskeiset kysymykset ja kohderyhmän. Ohjausryhmän kanssa keskusteltiin seminaaripaikasta ja ohjelman laatimisen periaatteista. Päätettiin pitäytymään 6.11.2019 ajankohdassa, vaikka seminaaritalan suhteen jouduttaisiin tekemään kompromissejä. Sovittiin, että osa seminaarinohjelmasta voidaan järjestää englanniksi.

minedu.fi/DL2021

Seminaarin lopulliseksi teemaksi tarkentui Datanhallinnan ja laskennan eurooppalaiset infrastruktuurit: Suomen näkökulma. Vuosiseminaarissa tarkasteltiin sitä, millaista kansallista ja Euroopan tason datanhallinnan ja laskennan ympäristöä (DL2021, EuroHPC, EOSC) Suomi haluaa olla rakentamassa.

7. Tiedoksi: Seuraavat kokoukset ja alustava listaus käsiteltävistä asioista

- 11.12.2019 klo 10-12, OKM > Kokous siirrettiin myöhemmin syksyllä pidettäväksi tammikuussa 2020.
- Kevät 2020: Sovittiin, että ohjausryhmälle lähetetään erillinen Doodle kokousajoista.
 - Maaliskuu (mm. DL2021-kehittämisohjelman tavoitteiden toteutumisen ja vaikuttavuuden seuranta)
 - Toukokuu/kesäkuun alku

Sovittiin, että pyritään järjestämään toinen ohjausryhmän kevään kokouksista Mahti-supertietokoneen avajaisten yhteydessä Kajaanissa.