

minedu.fi/DL2021

Datanhallinnan ja laskennan kehittämishjelman ohjausryhmä

Aika: Ma 14.12.2020 klo 10.00–12.00

Paikka: Etäkokous Teams-alustalla

Ohjausryhmän kokouksessa läsnä

- johtaja *Erja Heikkinen*, opetus- ja kulttuuriministeriö, pj.
- erityisasiantuntija *Teemupekka Virtanen*, sosiaali- ja terveysministeriö (saapui klo 10:30)
- neuvotteleva virkamies *Pirjo Kutinlahti*, työ- ja elinkeinoministeriö
- erityisasiantuntija *Lotta Engdahl*, liikenne- ja viestintäministeriö (neuvotteleva virkamies ~~Tuomas Kaivola~~, virkavapaalla 2020)
- professori *Sampsa Hautaniemi*, Helsingin yliopisto
- TKI-vararehtori *Kirsi Viskari*, Tampereen ammattikorkeakoulu
- ohjelmapäällikkö ~~Sirpa Thessler~~ (este), Luonnonvarakeskus (tietohallintojohtaja ~~Jukka Santala~~, Suomen ympäristökeskus)
- pääjohtaja *Heikki Mannila*, Suomen Akatemia
- Head of AI Business *Outi Keski-Äijö*, Business Finland Oy
- tutkimusjohtaja *Esa Virtanen*, ABB Power Grids Finland Oy (saapui klo 10:30)
- vararehtori *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 (asiantuntijasihteeri)
- johtava tiedeasiantuntija *Laura Taajamaa*, Suomen Akatemia (asiantuntijasihteeri)

Vierailijat

- professori *Keijo Heljanko*, Tieteellisen laskennan yhteistyöfoorumi, Helsingin yliopisto
- senior adviser *Johanna Törnroos*, CSC (saapui klo 10:10)
- johtaja *Pekka Lehtovuori*, CSC

minedu.fi/DL2021

Muistio

1. Kokouksen avaus ja asialistan hyväksyminen

Puheenjohtaja avasi etäkokouksen ja asialista hyväksyttiin.

2. Edellisen kokouksen muistio

Kokousmuistio ohjausryhmän 31.8.2020 pidetystä kokouksesta hyväksyttiin.

3. Tieteellisen laskennan yhteistyöfoorumin laatima viitearkkitehtuuri ja toimenpide-ehdotukset

Tieteellisen laskennan yhteistyöfoorumi (YTF) puheenjohtaja, Helsingin yliopiston rinnakkaisen ja hajautetun datatieteen professori Keijo Heljanko kertoi YTF:n luonnostelemasta [viitearkkitehtuuriraportista](#), jota käsiteltiin myös DL2021-ohjelman virtuaalisessa vuosiseminaarissa marraskuussa. Taustalla on isot investoinnit suurteholaskentaan sekä datanhallintaan ja raportissa on keskitytty pohtimaan, miten investointeja saadaan parhaalla mahdollisella tavalla hyödynnettyä: tarvitaan erityisosaamista muun muassa grafiikkasuorittimiin sekä tekoälyyn liittyen ja lisäksi käyttäjäpohjaa on laajennettava.

Heljanko kommentoi raportin kahdeksaa toimenpide-ehdotusta (kommentit kursivilla):

1. "On mahdollistettava ennustettava, pitkäaikainen tieteellisen laskennan infrastruktuurin rahoitus, joka mahdollistaisi tärkeiden suurinvestointien lisäksi tiheämmät infrastruktuurin päivityssykliä nopeasti muuttuvassa toimintaympäristössä."
Yhden, yhtenäisen laskentaympäristön mallin voi tulevaisuudessa korvata eriytyneemmät tarpeet.
2. "Tieteellisen datan koko elinkaaren aikaisen hallinnan viitearkkitehtuurin laatiminen tulee käynnistää. Tämän tulee sisältää myös suurten datamassojen pitkäaikaistallennuksen ja sille vaadittavan pitkän aikavälin rahoituksen suunnitelman. Suurten datamassojen pitkäaikaissäilytyksen vaatima infrastruktuuri avoimen tieteen periaatteiden mukaisesti on tärkeä tulevaisuuden kehityskohde. Lisäksi huomiota tulee kiinnittää sensitiiviseen dataan ja datan hallinnoinnin prosesseihin, mukaan lukien datan hyödyntämisen tutkimuseettiset kysymykset."
Tulee ratkaista, mikä data on säilytettävää ja miten tallennuksen kustannukset katetaan.
3. "Korkeakoulujen ja tutkimuslaitosten oman tieteellisen laskennan osaamisen parantaminen on tärkeää, esimerkiksi ammattimaisia Research Software Engineer ja Data Scientist -tehtäviä avaamalla tutkimuksen tueksi ja tieteellisen laskennan käyttäjäkunnan laajentamiseksi. Yksi tarvittavista toimenpiteistä on näissä rooleissa toimivien henkilöiden urapolkujen kehittäminen. Tällainen uusi tieteellisen laskennan osaajien ammatillista kehittymistä tukeva verkosto sekä tieteellisen laskennan palvelujen ja koulutuksen systemaattinen kehittäminen mahdollistavat vielä laajemman käyttäjäkunnan saamisen tieteellisen laskennan piiriin sekä tieteellisen laskennan palvelujen nykyistäkin tehokkaamman hyödyntämisen. Verkoston koordinoijana toimii CSC."
Laajeneva käyttäjäkunta tarvitsee räätälöityä tukea - pullonkaulana tällä hetkellä pieni osaajajoukko tarjoamaan paikallista käyttäjätukea sekä kansallisen koordinaation puute osaamiseen kehittämiseen. Tieteellisen laskennan asiantuntijaroolien kehityksen tulisi käynnistyä paikallisesti korkeakouluissa ja tutkimuslaitoksissa.

minedu.fi/DL2021

4. "Keskitetyn kansallisen infrastruktuurin palvelujen kehittämisen ohella myös tutkimuslaitosten ja korkeakoulujen oman laskentainfrastruktuurin kehittäminen erityisesti ohjelmistojen kehitysympäristöinä sekä suuria datamääriä tuottavien instrumenttien tiedon prosessointiin on tärkeää."
Erityisesti opiskelijatukea on hyvä tarjota ja sovelluskehitystä tehdä paikallisesti.
5. "Eurooppa-tason yhteistyötä tulee lisätä sekä infrastruktuurihankinnoissa että tieteellisen laskennan koulutuksessa ja sen tutkijoiden kansainvälisessä verkottamisessa."
On tärkeää edistää LUMlin linkittyvien tutkijoiden verkostoitumista.
6. "Tutkimuksen ja elinkeinoelämän yhteishankkeita mahdollistavia rahoitusohjelmia ja toimintatapoja tulee kehittää huomioimalla myös EuroHPC/Lumi-infrastruktuurin tuomat uudet mahdollisuudet. Business Finlandin roolia yrityspuolen TKI-rahoituksessa kehitetään."
Uusia toimintatapoja voidaan kehittää jo nyt, esimerkiksi datasettien jakamisen osalta.
7. "Vahvistetaan tutkimuspalveluiden laadun kehittämistä yhteistyössä asiakasorganisaatioiden kanssa hyödyntäen käyttäjälähtöistä palvelumuotoilua sekä dataa ja raportteja palveluiden käytöstä ja asiakastytytyväsyydestä."
Käytettävyyys ja sen seuraaminen on erittäin tärkeää käyttäjäkunnan laajentuessa. Tarvitaan verkkopohjaisia palveluita sekä uusia käyttöliittymiä ja ratkaistava miten räätälöidään erikoistuihin tarpeisiin.
8. "Tieteellisen ja dataintensiivisen laskennan kyvykkyyksien kehittämistä tukevia tutkimusohjelmia tulee avata Suomen Akatemian kautta."
Kilpailukykyisen infrastruktuuri-investoinnin täysimääräinen hyödyntäminen tarvitsee tuekseen tutkimusohjelmia.

Ohjausryhmä kiitti selkeästä ja innostavasta esityksestä sekä kokonaisarkkitehtuuriluonnoksesta.

Todettiin, että infrastruktuuri on nyt hyvällä tasolla, mutta osaamisen kehittämisessä sekä osaamiserojen kuromisessa on työtä. Koulutusta tulisi tarjota siellä, missä osaamista on. Myös koulutuksen kohdennus ja ketteryys sekä jatkuva oppiminen on tärkeää. Opiskelijamääriä ei voi merkittävästi kasvattaa samalla opetushenkilöstön määrällä.

Keskusteltiin korkeakoulujen staff scientist sekä data scientist -tyyppisistä urapoluista tutkimuksen ja hallinnon välimaastossa: esimerkkejä löytyy niin akatemian ja kliinisen puolen yhdistävästä polusta sairaalamaailmasta kuin bioinformatiikan puolelta. Pekka Lehtovuori kommentoi, että CSC tarjoaa yhden mahdollisen urapolun ja vetovoimatekijänä on mahdollisuus tukea laaja-alaisesti tutkimusta ainutlaatuisten infrastruktuurien parissa. Hän totesi myös, että laskentainfrastruktuurit pärjäsivät hyvin Akatemian FIRI-tiekarttahaussa. Osaajista kilpaillaan globaalisti ja työvoimapulaa on myös yrityksissä, joten tärkeää on miettiä myös houkuttelevuutta - monipuolisten työtehtävien tarjoaminen on kilpailuetu. Pohdittiin toisaalta myös monitieteisyyttä, miten geneerisempää dataosaamista ja toisaalta spesifimpää substanssipuolta painotetaan osaajien rekrytoinnissa. Samoin ajallinen pohdinta muutaman vuosikymmenen jänteellä olisi hyödyllinen urapolkuja kehitettäessä. Puheenjohtaja totesi, että tutkijanuratyöryhmä pohtii myös näitä kysymyksiä.

minedu.fi/DL2021

Laajenevan käyttäjäkunnan tarpeiden pohtimisessa ja paikallisen tuen kehittämisessä voisi hyödyntää palvelumuotoilua. Yritysyhteistyön uusien mahdollisuuksien miettimiseen voidaan saada tukea myös palvelumuotoilusta, mutta julkinen sektori (pl. Verohallinto) ei ole tässä vielä kovin pitkällä. Data-aineistojen jakaminen yritysten kanssa helpottuisi, mikäli jakaminen huomioitaisiin jo yhteistyösopimusten luonnosvaiheessa.

Ohjausryhmä keskusteli myös rahoituksesta: ennakoitava, pitkäjänteinen resursointi urapolkujen kehittämistyöhön paikalliseen tuen osaajien osalta mahdollistaa vaikuttavuuden loppukäyttäjille saakka. Resursseja kuluu janalla paikallisesta infrastruktuurista aina LUMlin saakka. Puheenjohtaja totesi, että rahoitusvirtoja tarvitaan useasta lähteestä. Puheenjohtaja totesi lisäksi, että YTF:n linjaukset ja tiedepolitiikkatyö sopivat hyvin yksiin ja yhteistyötä tarvitaan tässäkin.

Kokoustaako 5 minuuttia

4. DL2021-kehittämishjelma: Tavoitteiden toteutuminen ja vaikuttavuus – käyttäjänäkökulma

Johanna Törnroos CSC:ltä esitteli DL2021-ohjelman tavoitteiden toteutumista ja vaikuttavuutta käyttäjänäkökulmasta, taustalla on CSC:n käyttäjäkysely. Vastausprosentti on tyypillinen hieman työläämmälle kyselylle. Kyselyllä pyrittiin tavoittamaan myös ne, jotka eivät CSC:n palveluita vielä käytä, jotta voidaan kehittää toimintaa jatkossa. Kyselyyn vastanneet vastaavat profiililtaan ja tieteenalataustaltaan CSC:n käyttäjäkuntaa.

Palveluita käytetään pääasiassa tutkimukseen ja käyttäjinä pääasiassa opetus- ja tutkimushenkilöstö, mutta myös tukipalveluhenkilöstön osuus nousee (10 %). Vastaajat näkevät palveluiden tärkeimpinä hyötyinä tutkimuksen mahdollistamisen, tutkimusmetodien kehityksen ja kv-rahoitusmahdollisuuksien nostamisen. Voidaan todeta, että jos on osallistunut CSC:n koulutuksiin, palvelut koetaan käytettävämmiksi. Palveluita suositeltaisiin mielellään, Net Promoter Score (NPS-luku) on 68.

Osa kysymyksistä tarkasteli CSC:n palveluja paikallisen tuen ja infrastruktuurien jatkumona. Avovastauksista nousi esille se, että pääasiallisesti nämä täydentävät toisiaan, mutta aina paikallisia palveluita ei ole. Yhteistyötä taas tehdään paljon omalla sektorilla: ammattikorkeakoulut toisten ammattikorkeakoulujen kanssa, yliopistot ja tutkimuslaitokset pääasiassa yliopistojen kanssa, mutta myös kansainväliset tutkimuslaitokset nousevat vastauksissa.

DL-ympäristön eduksi aikaisempaan verrattuna vastaajat nostivat muun muassa Altaan (helpottaa datanhallintaa ja jakamista yhteistyökumppaneille ja toimii hyvin yhteen Puhdin ja Mahdin kanssa), nopeutuneen laskennan (mahdollistaa pidemmät, nopeammat ja toistettavat simulaatiot) sekä GPU-laskennan syvien neuroverkkojen mallien opetuksessa/laskennassa. Osa vastaajista ei ollut huomannut eroa aikaisempaan.

Vastaajat, jotka eivät ole käyttäneet palveluita totesivat, että eivät tiedä miten aloittaa, eivätkä osaa arvioida, mitkä palvelut sopisivat heidän tarpeisiinsa. Parannusehdotuksina palveluihin liittyen tai muuten CSC:lle esitettiin helppokäyttöisyyden edistämistä, aloittelijoiden ohjeistuksen sekä dokumentaation kehittämistä sekä jonojen lyhentämistä.

minedu.fi/DL2021

Ohjausryhmä kiitti esityksestä ja mielenkiintoisista tuloksista.

CSC:n toiminta ja palvelut ovat hyvällä mallilla ja on positiivinen tulos, että käyttäjät toteavat paikallisten ja CSC:n palveluiden toimivan hyvin yhteen. CSC jatkaa kyselyn tulosten tarkastelua muun muassa sen osalta, miksi 10 prosenttia vastanneista ja koulutuksiin osallistuneista ei edelleen löydä palveluista itselleen sopivaa. Pekka Lehtovuori totesi, että tämä tulos ei ole linjassa kurssikohtaisten palautekyselyjen kanssa.

Keskustelussa nostettiin esille, että koulutus ja tiedotus ovat isossa osassa vaikuttavuuden luomisessa: kenelle kohdistetaan tietoa ja millä tavalla. TKI-työn matala 20 prosentin osuus mietitytti, onko käyttäjille selvää, että palveluita voi käyttää yritysten kanssa tehtävässä yhteistyössä? Pekka Lehtovuori totesi, että kyselyä pitää myös kehittää, nyt ei ollut mahdollista valita useaa vastausvaihtoehtoa ja se saattaa näkyä myös tässä. Pekka Lehtovuori kiitti ohjausryhmää kommenteista ja muistutti, että opetuskäyttö avattiin vasta DL-ohjelman myötä ja on hyvä nähdä, että mahdollisuus on huomattu ja siihen tartuttu.

Outi Keski-Äijö kertoi Business Finlandin ja CSC:n uudesta [toimintamallista](#), jossa PK- sekä start up -yritykset voivat saada käyttöönsä enemmän laskentaresursseja osana AI Business -ohjelman projektiaan. Laskentatukea voi hakea tutkimus- ja kehitysrahoituksen yhteydessä max. 80 000 euroa, mikäli laskentakustannukset ylittävät 20 000 euroa.

Puheenjohtaja totesi lopuksi, että kysely osoittaa, että infrastruktuuri vastaa käyttäjien tarpeita ja kyselyn tuloksia on tarpeen hyödyntää myös jatkossa.

5. Vuoden 2021 kokousaikatauluehdotus ja muut tapahtumat

Sihteeristö esitteli vuoden 2021 suunnitelmat:

Kokous maaliskuussa

- LUMI-ekosysteemin käynnistyminen
- EOSC Forum Finland ja EOSC AISBL
- GAIA-X-hanke

[EuroHPC summit week](#) 22.-26.3.2020 (Virtuaalitapahtuma)

Kokous syyskuussa

- Tutkimukseen vaikuttava lainsäädäntö, erityisesti sensitiivinen data (toisiolaki, genomilaki, biopankkilaki)
- DL2021-kehittämishjelman vaikuttavuus
- DL2021-kehittämishjelman onnistumisen arviointia

minedu.fi/DL2021

DL2021-kehittämishjelman päätösseminaari tai -webinaari marraskuussa

- Järjestetään yhteistyössä CSC:n kanssa (ehdotus)
- Linkitetään tapahtuma myös Tutkitun tiedon teemavuoteen

6. Kokouksen päätös

Puheenjohtaja kiitti keskustelusta ja toivotti kaikille hyvää joulunaikaa. Kokous päättyi klo 11:52.