

minedu.fi/DL2021

Datanhallinnan ja laskennan kehittämishjelman ohjausryhmä

Aika: Ma 9.3.2020 klo 12.00–14.00

Paikka: nh. Aleksanteri, Kirkkokatu 12, Sisäministeriö

Osallistujat

johtaja *Erja Heikkinen*, opetus- ja kulttuuriministeriö, pj.
erityisasiantuntija *Teemupekka Virtanen*, sosiaali- ja terveysministeriö
neuvotteleva virkamies ~~*Pirjo Kutinlahti*~~, työ- ja elinkeinoministeriö
erityisasiantuntija ~~*Lotta Engdahl*~~, liikenne- ja viestintäministeriö (neuvotteleva virkamies ~~*Tuomas Kaivola*~~,
virkavapaalla 2020)
professori ~~*Sampsa Hautaniemi*~~, Helsingin yliopisto (yliopistotutkija *Susanna Pirttikangas*, Oulun yliopisto)
TKI-vararehtori *Kirsi Viskari (etäyhteydellä)*, Tampereen ammattikorkeakoulu
ohjelmapäällikkö *Sirpa Thessler*, Luonnonvarakeskus
pääjohtaja *Heikki Mannila*, Suomen Akatemia
Head of AI Business *Outi Keski-Äijö*, Business Finland Oy
tutkimusjohtaja *Esa Virtanen (etäyhteydellä)*, ABB Oy
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

johtaja *Pekka Lehtovuori*, CSC
senior adviser *Johanna Törnroos*, CSC

minedu.fi/DL2021

1. Kokouksen avaus ja asialistan hyväksyminen

Etäyhteydessä oli kuuluvuusongelmia, mutta täyden asialistan takia päätettiin kuitenkin aloittaa. Puheenjohtaja avasi kokouksen kello 12.08 ja asialista hyväksyttiin.

2. Edellisen kokouksen muistio

Hyväksyttiin kokousmuistio 20.1.2020 pidetystä ohjausryhmän kokouksesta.

3. DL2021-kehittämisohjelman tavoitteiden toteutumisen ja vaikuttavuuden seuranta

Etäyhteys saatiin toimimaan klo 12:15, joten sekä Kirsi Viskari että Esa Virtanen liittyivät kokoukseen tämän asiakohdan alkupäässä.

Johanna Törnroos esitteli CSC:n valmisteleman kaksiosaisen esityksen DL-ohjelman vaikuttavuuden seurannasta. Ensimmäisessä osassa hän kävi läpi tilastoja ja esimerkkejä ympäristön käytön ajalta ja toisessa osassa hän esitteli CSC:n arvion vaikuttavuuden mittareiden soveltuvuudesta. Esityksen aluksi Törnroos kertasi DL-ohjelman tavoitteet, summasi niiden toteutumiseen liittyvät tärkeimmät huomiot ja avasi esityksen rakennetta sekä tavoitteiden toteutumisen tietopohjaa sekä esimerkkitarkasteluja.

Infrastruktuuriin ja palveluihin liittyvät tavoitteet:

- Kehittämisohjelmalla päivitetään CSC – Tieteen tietotekniikan keskus Oy:n infrastruktuureja kansainvälisen yhteistyön varmistavalle tasolle.
- Uudistettava infrastruktuuri ja siihen liittyvät palvelut tarjotaan julkisen tutkimusyhteisön käyttöön aiempaa laajemmin. Valtion tutkimuslaitoksille tarjotaan vuodesta 2018 alkaen palvelut samoilla ehdoilla korkeakoulujen kanssa.
- Infrastruktuuri tukee tutkimusta ja opetusta.

Tutkimuslaitoskäyttäjien määrä CSC:n palveluissa kasvoi välillä 2018–2019 66 %, aktiivisten projektien määrä 143 %. Suurinta kasvu oli IL:ssä, LUKEssa ja VTT:llä, VTT on noussut jo 3. suurimmaksi (6. vuonna 2018). Tutkimuslaitosten projekteista 29 % on yhteisprojekteja yliopiston tai AMK:n kanssa. Infrastruktuuri hyödytti jo pilottivaiheessaan niin suomalaisten tutkijoiden omaa tutkimusta kuin ja yhteistyötä kansainvälisten ryhmien kanssa. Suurimmissa tieteenaloissa resurssien käytöllä mitattuna tapahtui vuosina 2018–2019 muutoksia, kun biotieteet kasvoi fysiikan ohi suurimmaksi tieteenalaksi (biotieteet lähes 30 % resursseista, fysiikka 23 %) ja tietojenkäsittely- ja informaatiotieteet ja avaruus- ja tähtitieteet nousivat nanoteknologian edelle.

Osaamiseen liittyvät tavoitteet:

- Huomioidaan uudentyyppiset ja nousevat tutkimusyhteisön palvelutarpeet sekä uudet tutkimusalat.
- Mahdollistetaan data-aineistojen ja laskennan hallinnointi siten, että tutkimus- ja innovaatiotoiminta pystyy vastaamaan myös uusiin haasteisiin.
- Vahvistetaan kansallista ja helpotetaan paikallista asiantuntijatukea keinoilla, jotka skaalautuvat kasvavan käyttäjämäärän ja tutkimuslaitosten mukaantulon myötä.

Useilla tieteenaloilla on kasvua resurssien käytössä ja aktiivisten projektien määrässä. Tekoälytutkimus näkyy uuden infrastruktuurin alkuvaiheen projekteissa, noin 17 % projekteista linkittyy tekoälyyn. FCAI-CSC

minedu.fi/DL2021

yhteistyöllä edistetään tekoälylaskennan ja tekoälymenetelmien hyödyntämistä. VTT:n projektien määrän kasvu ja FCAI-CSC yhteistyö indikoivat ohjelman hyödyistä myös elinkeinoelämälle. Lisääntyvien webinaarien myötä osaamisen kehittämistä saadaan skaalattua kasvavan asiakaskunnan tarpeisiin.

Ohjausryhmä kiitti Johanna ja Pekkaa hyvin valmistellusta esityksestä ja keskusteli nousevista uusista aloista, biotieteet kasvaneet fysiikan ohi tieteenalakohtaisessa tarkastelussa muuttuneen tilastointitavan myötä. Uudella tilastointitavalla tehtiin näkyväksi biotieteilijöiden käyttö, nyt mittautapa on yhteismitallinen. Myös kielitieteet on noussut uutena tieteenalana Puhdin pilottiprojektin myötä. Ohjausryhmä keskusteli myös infrastruktuurista korkeatasoisen tutkimuksen mahdollistajana. Esimerkkeinä käytettyjä tutkijoita on rahoitettu ERC:n lisäksi myös kansallisesti, esimerkiksi Suomen Akatemian PROFI-rahoituksen kautta.

Pekka Lehtovuori selvensi projektien hyväksymisprosenttia: projektihakemuksia ei hylätä, mutta laskentaresursseja saatetaan myöntää vähemmän kuin haettu. Pekka kertoi myös, että pilottien tuloksia ja esimerkiksi julkaisutoimintaa seurataan CSC:n toimesta. Ohjausryhmän kysymykseen, että olisiko osa projekteista ollut toteutettavissa vanhassa ympäristössä, Pekka kommentoi, että esimerkiksi tekoälyhankkeita ei olisi voitu toteuttaa. Todettiin, että ammattikorkeakoulujen käyttäjämäärä kasvanut, vaikka absoluuttisesti määrät edelleen pieniä. Pekka Lehtovuori kertoi, että ammattikorkeakoulut ovat ottaneet aktiiviseen käyttöön muun muassa pilvikapasiteettia sekä konttipilvipalveluita. Resurssihakemuksista ei nykyisin suoraan näy onko suunnitellussa hankkeessa mukana yrityksiä. Ohjausryhmä keskusteli siitä, voisiko nykyistä hakulomaketta räätälöidä siten, että mahdollistettaisiin nykyistä helpommin tiedolla johtaminen. Yhteistyö ja sen tarkempi ymmärtäminen olisi kiintoisaa.

Esityksen toisessa osassa Johanna Törnroos kävi läpi ohjelman vaikuttavuuteen suunniteltua mittaristoa. Edellisen kerran ohjausryhmä on keskustellut ohjelman vaikuttavuudesta vuoden 2008 lopulla ja CSC on nyt tarkastellut tuolloin luonnosteltujen mittareiden soveltuvuutta. Mittareiden arvioinnin tarkoitus varmistaa tarkoituksenmukainen raportointi. Osa vuonna 2018 määritellyistä mittareista on oleellisia ohjelman tavoitteiden toteutumisen sekä vaikuttavuuden seurannan kannalta ja osa soveltuu ohjelman vaikuttavuuden tarkasteluun pidemmällä aikavälillä, kun taas osa mittareista on oleellisia tarkasteltaessa CSC:n roolia Suomen tki-kentässä, mutta eivät välttämättä ole tarkoituksenmukaisia DL2021-ohjelman vaikuttavuuden tarkastelussa.

- Lyhyen aikavälin seuranta
 - o Tutkimuslaitosten asiakassegmentin kehittyminen
 - o Uusien tutkimusalojen käyttäjämäärien ja käytön kehittyminen
 - o Tekoälytutkimuksen ja -menetelmien käytön kehittyminen
 - o Uusien palvelujen käytön kehittyminen
- Keskipitkän aikavälin seuranta
 - o Infrastruktuurin merkitys tutkimukselle ja opetukselle (esim. kv-yhteistyö, läpimurrot, menetelmien kehitys)
 - o Asiakkaiden kyvykkyyksien kehittyminen palvelujen käyttämisessä
 - o Infrastruktuurin merkitys elinkeinoelämälle ja innovaatioille
- Pidemmän aikavälin seuranta
 - o Tieteen tilan kehittyminen laskennallisilla ja dataintensiivisillä tieteenaloilla (julkaisuanalyysi)
 - o Yhteiskunnallisesti merkittävien tutkimustulosten nosto ja kansantajuinen viestintä (myös keskipitkällä ja lyhyellä aikavälillä)

minedu.fi/DL2021

Ohjausryhmä keskusteli koulutusten tärkeydestä, ja erityisesti notebooks-palvelu nostettiin esiin. Tietty ammattikorkeakoulut ovat myös suunnitelleet CSC-ajokorttia, olisiko osallistuja- ja käyttäjämääriä mahdollista seurata tätä kautta. Pekka Lehtovuori totesi, että opetuskäyttö haastaa hallitsemaan odotuksia ja myös käyttäjätukea erilaisen käyttäjäprofiilin kautta. Puheenjohtaja summasi, että odotusarvona ei pidetä sitä, että kaikki olisi heti valmista, vaan ongelmia ratkotaan, kun niitä ilmenee. Lopuksi puheenjohtaja kiitti vielä CSC:tä esityksestä.

4. Katsaus tutkimuslaitosten näkemyksiin DL-palveluista

Sirpa Thessler esitteli tutkimuslaitosten huomioita DL-ympäristön käyttöönotosta. Uuden laskentaympäristön laventaminen myös tutkimuslaitosten käyttöön on merkittävä parannus ja käyttö on kasvanut. Tarve tutkimuslaitosten omaan kapasiteettiin ei kuitenkaan ole kokonaan poistunut. THL huomioi, että arkaluontoisen datan käsittely on korostunut ohjelman myötä. Tekoäly on tulossa tutkimuslaitosten operatiiviseen tiedontuotantoon, ja sitä ei voi toteuttaa DL-ympäristössä. Keskustelussaan ohjausryhmä totesi, että tekoälyn opetusvaiheen jälkeen DL-ympäristön tarve ei ole välttämätön. Tutkimuslaitokset nostivat myös esille, että EuroHPC:n voisi hyödyntää tutkimuslaitosten tutkimuksessa ja yritysyhteistyössä, samoin yhteistyötä eurooppalaisten tutkimusinfrastruktuurien kanssa voisi tiivistää. Yhteinen kehitysympäristö nopeuttaisi analysoinnin aloitusta ja pidemmälle vietyjen alakohtaisten tekniikoiden hyödyntämistä.

Tutkimuslaitoksilla on lakisääteinen roolitus ja omanlaisensa tarpeet roolien myötä – esimerkiksi jatkuvaluonteiselle seurannalle ja pitkäaikaissäilytykselle. Yhä useammin hyödynnetään ja tuotetaan isoja tutkimusaineistoja, joten CSC:n ja Valtorin yhteistyön toimivuus korostuu. Koska tutkimuslaitokset siirtävät isojakin datamassoja tutkimusyhteisön ja julkisen hallinnon järjestelmien välillä, tiedonsiirto ei saisi muodostua pullonkaulaksi. Tutkimuslaitoksista oli noussut esiin myös kaupallisten pilvipalvelujen mahdollinen rooli esimerkiksi Valtorin palveluna.

Käyttäjät/tutkijat ovat tyytyväisiä CSC:n laskentapalveluihin. Lisenssiehdoissa olisi hyvä mahdollisuuksien mukaan huomioida myös tutkimuslaitoksille tyypillinen käyttö kuten asiakasrahoitteiset hankkeet ja ministeriöiden toimeksiannot. Toiminnan kehittämiskohteena tutkimuslaitokset näkevät ympäristön riittävän skaalautuminen projektien tarpeisiin, esimerkiksi nopean väylän levykapasiteetin kasvattaminen kymmenien/satojen teratavujen kokoluokkaan. CSC:n järjestämät koulutukset on koettu hyödylliseksi ja sitä pidetään tärkeänä, että koulutusta tarjotaan keskitetysti. Käyttäjätuesta on tullut positiivista palautetta.

Ohjausryhmä pohti julkisen sektorin tuottaman tiedon hyödyntämistä yrityksissä ja CSC:n mahdollisuutta toimia sillonrakentajana, jotta yhteiskunnan kannalta arvokasta tietoa saataisiin käyttöön. Thessler huomautti, että tutkimuslaitoksissa tehdään yritysyhteistyötä muun muassa osaamisen saralla. Ohjausryhmä keskusteli mahdollisista kehityskohteista palveluissa, erityisesti tutkimuslaitosten näkökulmasta. Todettiin, että haasteita ratkotaan yhteistyönä esimerkiksi Valtorin kanssa. Palveluiden maksuttomuus/maksullisuus on tullut esille myös yhteydenotoissa CSC:lle, mutta asia on osa laajempaa kokonaisuutta eikä myöskään yksin CSC:n päätettävissä.

Lopuksi ohjausryhmä kiitti Sirpa Thessleriä tärkeästä koosteesta ja palautteesta sekä totesi, että on mielenkiintoista kuulla käyttötapauksista. On hyvä kuulla, että asiat ovat edenneet ja että haasteista ratkotaan yhdessä. Puheenjohtaja totesi, että CSC:n ja Valtorin yhteistyöhön palataan tämän vuoden aikana ohjausryhmän kokouksessa. Kehittämisohjelman aikana ratkotaan osa tässäkin esityksessä nousseista haasteista OKM:n voimin ja joidenkin osalta on hyvä tutkimuslaitosten keskustella omien ministeriöiden

minedu.fi/DL2021

kanssa. Osaa voidaan helpottaa viestinnällisin keinoin CSC:n kanssa yhdessä. Keskusteluja tutkimuslaitosten omien ministeriöiden kanssa helpottaa, jos on selkeästi todettu, mikä on maksutonta käyttöä ja mikä olisi mahdollista maksua vastaan. Pekka Lehtovuori kommentoi, että esimerkiksi Ilmatieteenlaitoksen oma kapasiteetti on hankittu Puhdin yhteyteen, jolloin on saatu synergiaetua ylläpidossa ja parannetaan loppukäyttäjän kokemuksia. Keskustelussa nousi esille, että myös korkeakouluilla on maksullista käyttöä. Ohjausryhmä keskusteli myös siitä, että mikäli käyttäjien tarpeet laajenevat, miten CSC:n kapasiteettia priorisoidaan. Puheenjohtaja totesi, että priorisoinnissa apuna on niin CSC:n neuvottelukunta kuin hallituskin sekä YTF:n kautta tutkijat.

Erja Heikkinen poistui kokouksesta tämän asiakohdan käsittelyn päätteeksi klo 13:13 ja puheenjohtajana toimi kokouksen loppuun saakka ohjausryhmän varapuheenjohtaja Heikki Mannila.

5. EOSC:in valmistelussa ajankohtaista

Juha Haataja kertoi ohjausryhmälle EOSCin ajankohtaisista käänteistä. EOSC on nostettu esille juuri julkaistussa Euroopan datastrategiassa ja tämä tulee ohjaamaan myös kansallisella tasolla. Komissio linjasi, että EOSC on tuotannossa vuonna 2025. Valmistelussa on tällä hetkellä Belgian mallinen AISBL-yhdistys ja tavoitteena on perustaa oikeushenkilö vuoden 2020 aikana. Suomi seuraa valmisteluja: miten nykyiset valmistelevat ryhmät suhtautuvat uusiin ja miten maat sekä ryhmät asemoituvat yhdistyksen sisällä. Kysymyksenä ohjausryhmän keskustelun pohjaksi Haataja nosti sen, että mikä taho voisi liittyä Suomesta EOSC-yhdistyksen ensijäseniksi.

Ohjausryhmä pohti mikä muuttuu, mitä ollaan ratkaisemassa ja miten kansalliset palaset liittyvät tähän eurooppalaiseen kokonaisuuteen. Valmistelua on leimannut keskustelu erilaisten visioiden ja näkökantojen sovittelusta, mutta yleisenä tavoitteena on se, että olemassa olevat palvelut tutkijoille tehdään yhteensopiviksi, jotta tiedonsiirto helpottuu. EOSC tarjoaa mahdollisuuden avata kansallisia palveluja laajempaan käyttöön ja saada vastinrahaa komissiolta. Pekka Lehtovuori kertoi pilotoitavasta infra-EOSC -hausta, jossa komissio kattaa palveluntarjoajan kustannuksia uusien virtuaalikäyttäjien osalta. Keskustelussa nousi esille myös yrityksille suunnattu eurooppalainen GAIA-X -projekti datainfrastruktuurien federointiin.

Ohjausryhmä keskusteli myös EOSC-hallinnosta ja sen mahdollisista riskeistä: yhdistyksellä mahdollisilla jäsenillä on toisaalta rooli politiikan muodostajana ja toisaalta mahdollisena rahoituksen saajana. Haataja kertoi, että luonnosteltava oikeushenkilömalli ratkonee osan näistä haasteista. Ohjausryhmä toivoi, että Suomi veisi aktiivisesti viestiä käyttäjän/tutkimuksen tekijöiden näkökulman huomioimisesta sekä siitä miten kansalliset palvelut kytkeytyvät kokonaisuuteen. Ohjausryhmä toivoi myös, että valmisteluun tulisi konkretiaa mahdollisimman pian. Haataja totesi, että suomalaiset ovat olleet valmistelussa aktiivisesti mukana ja myös koordinoivat toimia keskenään. Suomella on hyvät lähtökohdat, EOSCin tavoitteena osaltaan on se, että Euroopassa päästään samalle tasolle.

6. EuroHPC: Ajankohtaiset asiat

Pekka Lehtovuori kertoi EuroHPC:n ajankohtaisista asioista. Palvelusopimus ja rahoitussopimus solmittu EuroHPC Joint Undertakingin kanssa ja LUMI-konsortiosopimus on valmisteilla. Islannista on tullut EuroHPC-yhteisyrityksen 32. jäsenmaa ja CSC käy keskusteluja Islannin mahdollisesta liittymisestä LUMI-konsortioon. LUMI:n datakeskuksen valmistelut myöhästyvät muutaman kuukauden suunnitellusta. LUMI-laitteiston kilpailutus edennyt toiseen vaiheeseen eli neuvotteluihin laitetoimittajien kanssa, alustavia tarjouksia odotetaan 30.3. LUMI-ohjelman kansalliset käyttäjäohjelmat on luonnosteltu ja myös resurssiallokaation

minedu.fi/DL2021

periaatteet ovat työn alla. Petteri Kauppinen lisäsi, että EuroHPC-yhteisyrityksen puolella katsotaan kansainvälistä käyttäjäpolitiikkaa ja se on tarkoitus saada valmiiksi vuoden 2020 aikana.

Lehtovuori kertoi myös aiheeseen liittyvistä osaamiskeskuksista. Keskuksia on suunnitteilla yksi per maa, ja ne sijoitetaan supertietokonekeskuksiin. Keskus tarjoaa mm. pääsyn EuroHPC-infrastruktuureihin, mahdollisuuden asiantuntijakonsultaatioon, koulutusta ja viestintää sekä verkostoitumista muiden keskusten kanssa. Toiminta alkamassa keväällä 2020, ensimmäinen kausi kaksivuotinen. Suomen osaamiskeskus keskittyy LUMIn "L3"-tukeen suurteholaskennan, data-analytiikan ja tekoälyn menetelmien osalta. Ohjausryhmä totesi, että komission 1 MEUR/2v panostus 1 MEUR kansallisine vastinrahoineen osaamiskeskuksen kehittämiseen on merkittävä.

Lisäksi Lehtovuori jakoi ohjausryhmälle "LUMI User Program – Finland" -tiedoston ja pyysi lähettämään mahdolliset kommentit hänelle sähköpostitse. Ohjausryhmä pohti mikä taho olisi oikea ottamaan tähän kantaa.

7. Tiedoksi: Seuraavat kokoukset ja alustava listaus käsiteltävistä asioista

Puheenjohtaja muistutti ohjausryhmän jäseniä hoitamaan Kajaanin kokouksen lentolippujen hankinnan ajoissa. Etäosallistumismahdollisuuden järjestäminen on tärkeää, koska samaan aikaan järjestetään valtakunnalliset ammattikorkeakoulupäivät. [Ohjausryhmän kokousaikataulu vuodelle 2020 on päivittynyt ja tarkentunut tämän kokouksen jälkeen.]

5.5. Kokous Kajaanissa Mahti-superkoneen avajaisten yhteydessä - alustava ohjelma

- 8.10-9.25 Menolento
 - **10.00-12.00 DL2021-ohjausryhmän kokous (Kongressi- ja kulttuurikeskus Kaukametsä, Koskikatu 2-4)**
 - o Kansallinen TKI-tiekartta
 - o EuroHPC kansallinen käyttäjäpolitiikka
 - o Korkeakoulujen DL-näkökulma
 - o CSC:n oma strategiatyö DL-näkökulmasta
 - 12.00-12.30 Lounas
 - 12.30-16.30 Mahdin avajaisjuhlallisuudet
 - 18.10-19.25 Paluulentomahdollisuus
- (Illallinen Kajaanissa, CSC järjestää osana avajaisjuhlallisuuksia.)

Syksyn kokoukset, käsiteltäviä aiheita

- Tieteellisen laskennan yhteistyöfoorum (YTF) kokonaisarkkitehtuurityö
- Tutkimusdatan avoimen saatavuuden linjauksen valmistelu TSV:n koordinoimana
- DL2021-kehittämishjelmaan linkittyvät lainsäädännölliset asiat (jatkoa aikaisempiin kokouksiin: Toisiolaki, Genomilaki, Biopankkilaki)

Vuosiseminaari (alustavasti marraskuu)

- Alustavasti suunniteltu toteutusta yhteistyössä YTF:n kanssa
- YTF:n arkkitehtuurityö loppusuoralla (arkkitehtuurissa huomioidaan vahvasti kansainvälinen kehitys)
- LUMI tulossa asiakaskäyttöön 2021