



## Tutkimuksen digitalisaation politiikkatoimien kokoamisen ja kansallisen datanhallinnan ja laskennan tutkimusinfrastruktuuriympäristön päivittämisen (DL2026) seurantaryhmä

**Aika:** keskiviikko 10.9.2025 klo 12.30–14.00

**Paikka:** sisäministeriö, kh. Maria, Kirkkokatu 12

### Osallistujat:

johtaja Erja Heikkinen, opetus- ja kulttuuriministeriö, puheenjohtaja  
tutkimusjohtaja Annukka Pakarinen, ammattikorkeakoulujen rehtorineuvosto Arene ry (Hämeen ammattikorkeakoulu)  
~~yksikön johtaja Niko-Matti Ronkonmäki, liikenne- ja viestintäministeriö (este)~~  
johtava tiedeasiantuntija Risto Vilkkö, Suomen Akatemia  
tutkimuspalveluiden johtaja Ella Bingham, Suomen yliopistojen rehtorineuvosto Unifi ry (Aalto-yliopisto)  
~~professori Minna Palmroth, Suomen Tiedeakatemia (Helsingin yliopisto) (este)~~  
~~johtaja Pekka Lehtovuori, Tieteen tietotekniikan keskus CSC (este)~~  
~~yksikönjohtaja Saku Anttila, Tutkimuslaitosten yhteenliittymä Tulanet (Suomen ympäristökeskus) (este)~~  
opetusneuvos Petteri Kauppinen, opetus- ja kulttuuriministeriö  
opetusneuvos Laura Taajamaa, sihteeri  
  
Pysyvät asiantuntijat  
professori Heikki Mannila, Aalto-yliopisto  
CSC:n hallituksen puheenjohtaja Jukka Mönkkönen, Itä-Suomen yliopisto

asiakasratkaisupäällikkö Outi Tasala, Tieteen tietotekniikan keskus CSC (asiakohdat 1-3)  
ryhmäpäällikkö Eero Siivola, Suomen ympäristökeskus

### 1. Kokouksen avaus ja asialistan hyväksyminen

Puheenjohtaja avasi kokouksen klo 12.30. Ryhmään toivotettiin tervetulleeksi uusi jäsen Risto Vilkkö. Saku Anttilan sijaisena tässä kokouksessa on ryhmäpäällikkö Eero Siivola. Asialista hyväksyttiin.

### 2. Edellisen kokouksen muistion hyväksyminen

Edellisen kokouksen ja vuosiseminaarin muistio hyväksyttiin. Helmikuun muistiossa havaittiin pieniä muotoiluvirheitä, jotka korjataan ja päivitetään Tiimeriin.

### 3. Tilannekatsaus LUMI-AI-supertietokoneen hankintaan ja toiminnan käynnistämiseen

Outi Tasala (CSC) esitteli hankinnan ja LUMI-tekoälytehtaan ajankohtaisen tilanteen. Tavoitteena tehtaassa on mahdollistaa tutkijoille helppo tapa saada tekoälymenetelmät käyttöön. Tehdas rakentuu kolmesta pääkomponentista: tekoälyoptimoidusta supertietokoneesta (LUMI-AI), käyttäjiä tukevasta palvelukeskuksesta sekä kokeellisesta kvanttilaskenta-alustasta (LUMI-IQ). Tehdasta koordinoi Suomi (CSC) ja mukana kansainvälisessä konsortiossa ovat Tšekki, Tanska, Viro, Norja ja Puola. EuroHPC-yhteisyrityksen antennihaussa kiinnostuksensa ovat osoittaneet myös Sveitsi, Belgia, Islanti ja Latvia. Kansallisesti mukana on lippulaiva FCAI ja AI Finland. Kokonaisbudjetti on yli 612 milj. euroa, josta EU:n osuus on 306,4 milj. euroa.

Uuden LUMI-AI-superkoneen konesalin pääurakoitsija valittiin elokuussa 2025 ja rakennusvaihe alkaa vuoden vaihteessa 2025–2026. Konesali on aikataulun mukaan valmiina supertietokoneen asennukselle kesällä 2027. LUMI-IQ:n osalta hankintapyyntö julkaistaan lokakuussa 2025 ja asennus on suunniteltu vuoden 2027 toiselle neljännekselle. Ohjelmistot kehittyvät vaiheittain vuosina 2025–2032.

LUMI-tekoälytehdas avattiin nykyisen LUMI-supertietokoneen yhteyteen huhtikuussa ja verkkosivusto julkaistaan syyskuussa. Ensimmäiset palvelut sisältävät laskentaa eri kokoluokissa, tukea resurssihakemuksiin ja teknisiin arvioihin, konsultointia ja optimointia AI/HPC-projekteihin, DaaS eli Datasets-as-a-Service-käyttäjätukea sekä hajautetun mallikoulutuksen asiantuntijapalveluita. Ensimmäiset asiakkuudet ja koulutukset, kuten kvanttimenetelmiin liittyvät kurssit, ovat jo käynnistyneet. Lisäksi terveysalan fokusryhmä on aloittanut toimintansa.

Puheenjohtaja kiitti selkeästä esityksestä. Todettiin, että tietoa datakeskuksen rakennuttajasta ei vielä voi jakaa. Kauppinen kysyi tekoälytehtaiden yhteistyöstä ja Tasala totesi, että yhteistyöhön kannustetaan EuroHPC-yhteisyrityksen taholta. CSC:llä on toimia aiheen tiimoilta ja etuna on se, että toiminta saatiin käyntiin nopeasti nykyisen LUMI:n yhteyteen. EuroHPC-yhteisyrityksellä on käynnissä haku, jolla etsitään toimijaa tehtaiden välille. Todettiin myös, että vaikka monia hankintoja on käynnissä samanaikaisesti, ne etenevät aikataulussa.

Lisäksi keskusteltiin palveluosion oleellisesta Datasets-as-a-service -toiminnosta ja siitä, mistä erilaisia datasettejä on saatavilla. Palvelussa on tunnistettu ja hyödynnetään olemassa olevia settejä (mm. kielipankki). Myös konsortiopartnereiden sekä asiakkaiden myötä laajennetaan valikoimaa. Pohdittiin myös miten työ nivoutuu kansalliseen yhteistyöhön datan hyödyntämisessä esimerkiksi Tilastokeskus tai Findata. Tähän kysymykseen palataan.

#### **4. Yliopistojen tutkimusdatan hallinnan ja säilytyksen selvitystyö**

Jukka Mönkkönen esitteli ehdotusta yliopistojen tutkimusdatan hallinnan kansalliseksi kehittämismalliksi. Opetus- ja kulttuuriministeriön toimeksiannosta toteutettu selvitystyö kartoittaa yliopistojen tarpeita ja tahtotilaa tutkimusdatan hallinnan palveluiden kansalliselle järjestämiselle. Tavoitteena parantaa tutkijoiden arkea käytännössä. Selvityksen taustalla toimii DAHA-viitearkkitehtuuri, ja työtä on ohjannut Tieteellisen laskennan ja datanhallinnan yhteistyöfoorumi (YTF). Työryhmään on osallistunut asiantuntijoita eri yliopistoista sekä CSC:stä.

Selvityksessä havaittiin, että tutkijoiden käytössä on kattava, mutta hajautunut palveluvalikoima. Palvelupolku ei ole selkeä ja palveluiden tunnettuudessa on parantamisen varaa. Selvityksen mukaan tarvitaan helposti hahmotettava palvelukokonaisuus, joka tukee datan elinkaaren hallintaa keruusta julkaisuun. Keskeiseksi ratkaisuksi esitetään Dataset-as-a-Service (DaaS) -mallia, jossa tutkijoille tarjotaan valmiita, kuratoituja aineistoja tietoturvallisessa ympäristössä. Palveluun sisältyisi myös luvitus arkaluonteisen datan käsittelyyn. Kansalliset palvelut, kuten Allas, Pouta/ePouta, Fairdata, LUMI ja EUDAT,

tukevat eri vaiheita datan hallinnassa ja mahdollistavat tehokkaan tutkimustyön. Tutkimusinfrastruktuurit, kuten Tietoarkisto, FIN-EPOS ja INAR RI, tarjoavat alakohtaisia palveluita, jotka tulee myös huomioida kansallisessa kokonaisuudessa. Näiden infrastruktuurien rooli datan tuottamisessa, analysoinnissa ja jakamisessa on merkittävä.

Selvityksessä ehdotetaan seuraavia toimenpiteitä

1. **Toiminnallinen tutkimusdatanhallinnan suunnittelu ja työkalut sen tukemiseen:** Organisaatiot vaativat ja tukevat tutkijoita tekemään ja päivittämään datanhallintasuunnitelmat. Organisaatiot kasvattavat omia kyvykkyyksiään hyödyntää datanhallintasuunnitelmia palveluiden resursoinnissa, kohdentamisessa ja kehittämisessä. Toiminnallisen (koneluettavan) datanhallintasuunnitelmatyökalun kehittäminen tukemaan datanhallinnan prosessia ja toimintaa.
  - ➔ Toimet helpottavat tutkijan arkea ja tutkimustyö sekä meritoituminen datanhallinnasta tehostuu. Organisaation datanhallinnan tarpeiden ennustattavuus paranee ja arviointi mahdollistuu.
2. **Palveluiden yhteentoimivuuden varmistaminen:** aito yhteistyö paikallisten ja kansallisten toimijoiden kesken vahvistuu: selkiytetään työnjakoa ja jaetaan parhaita käytäntöjä. Rakennetaan kansallinen datanhallinnan palvelukokonaisuus ja datanhallinnan toimintamallit ja prosessit, jotka täyttävät kansainväliset yhteentoimivuuden vaatimukset ja standardit, jolloin datan siirtäminen palveluiden välillä on helppoa. Kokonaisuus mahdollistaa paikallisten palveluiden joustavan ja jatkuvan kehittämisen osaksi laajempia kansallisia ratkaisuja.
  - ➔ Tutkijat voivat hyödyntää palveluja paikallisesta kansainväliseen ja datanhallinta on sujuvaa eri ympäristöissä.
3. **Arkaluonteisen datan palveluiden jatkuva kehittäminen:** luottamuksellisen datan palveluiden jatkuva kehittäminen paikallisesti ja kansallisesti ml. tehokkaimmat supertietokoneet. Kansallisten SD (sensitive data) -palveluiden laajempi hyödyntäminen. Osaamisen ja tuen varmistaminen. Selvitetään kansallisen anonymisointipalvelun kehittämistä ja laaditaan kryptausteknologioita koskeva ohjeistus. Sektorirajat ylittävää yhteistyötä tarvitaan. Selvityksessä rajattu sote-datan käyttö ulkopuolelle.
  - ➔ Arkaluonteisten aineistojen käyttö tutkimuksessa tehostuu ja monipuolistuu.
4. **Pitkäaikaissäilytyspalvelun käytön laajentaminen:** aineistojen kuvailun tehostaminen eli metatietojen tuottaminen mahdollisimman varhaisessa vaiheessa. Organisaation sisäisten prosessien kehittäminen määrittelemällä datanhallinnan peruspalvelut sekä laatimalla organisaation datastrategian. Luodaan kansallinen ohjeistus tutkimusaineistojen arvonmäärittelyyn.

- ➔ Arvokkaat tutkimusaineistot säilyvät laadukkaina ja käytettävänä jälkipolville asti. Aineistojen autenttisuus voidaan taata. Aineistojen hallinnan merkittävä parantuminen. Organisaatioiden oman tallennusinfra keventyminen.

5. **Tallennuskapasiteetin riittävyys laajoille tutkimusaineistoille:** Kehitetään palveluiden soveltuvuutta eri tutkimusalojen ja erilaisten datojen käsittelyn näkökulmasta ja huomioidaan erityisesti paljon tallennustilaa vaativat datat. Tämä edellyttää aineistojen elinkaaren hallintaa (ml. datan hallittu hävittäminen). Sovitaan työnjaosta, organisaatiokohtaiset ja kansalliset ratkaisut. Kehitetään yhteentoimivuutta. Palvelujen tarjoajien kesken sovitaan uudenaikaisesta kehittämisestä toimintamallista, jota vaikuttavuusperusteinen strateginen kehittäminen edellyttää.

- ➔ Tutkimusaineistot saadaan talteen, tutkimuksen tekeminen tehostuu

6. **Strategisen johtamisen ja toiminnan toteuttamisen välisen yhteyden tiivistäminen:** Kansallisella ja paikallisella tasolla tarvitaan yhteinen datanhallinnan tavoitetilä<sup>1</sup>. Organisaation johdon ja palveluiden (datatuki, kirjastot, IT-palvelut, tutkimuspalvelut) välisen yhteistyön vahvistaminen ja kansallisten verkostojen ja toimielinten (esim. YTF) roolien selkiyttäminen ja vahvistaminen tärkeää. Toimijoiden sitouttaminen jaettuun tavoitetilään mahdollistaa systeemisen tai verkostomaisen johtamisen.

- ➔ Yhteentoimivien palveluiden tuottaminen tehokkaasti yli hallinto- ja organisaatorajojen.

7. **Kansallinen palvelukatalogi:** kootaan kansallisesti saatavilla olevat palvelut samaan katalogiin, luodaan korkeakoulujen omat palvelukatalogit ja CSC:n palvelukatalogi integroidaan paikallisiin palvelukatalogeihin ja -näkyymiin. Koneluettavuus ja linkittyminen koneluettavaan datanhallintasuunnitelmaan oleellista.

- ➔ Palveluiden löydettävyys parantuu ja mahdollistaa soveltuvien palvelujen automatisoidun tarjoamisen tutkimuksen elinkaaren vaiheisiin (tutkijan palvelupolut)

Esityksen lopuksi Mönkkönen korosti, että työ koskee yhtä lailla ammattikorkeakouluja ja tutkimuslaitoksia ja ne ovat myös YTF:ssä edustettuna.

Seurantaryhmä kiitti kattavasta esityksestä ja todettiin, että pitkäaikaissäilytyksen sijaan olisi parempi käyttää termiä pitkäaikaissaavutettavuus. Heikki Mannila kysyi, oliko selvityksen aikana keskusteltu tarvittavasta henkilöstörakenteesta ja -profiileista? Mönkkönen totesi, että tässä ensimmäisessä vaiheessa selvitettiin mitä palveluita on, seuraava askel on tarkastella mitä vielä tarvitaan ja kaikkea tietoa ei välttämättä ole saatavilla.

---

<sup>1</sup> <https://wiki.eduuni.fi/spaces/CscKtpoYtf/pages/494637597/3.4+Palvelut#id-3.4Palvelut-3.4.2.5Tutkimusendatanhallinnanjohtaminenjaorganisointi>

Annukka Pakarinen totesi, että asia on ajankohtainen ammattikorkeakouluissa ja erilaisia keskustelunavauksia on ja myös CSC:n kanssa keskustellaan. Mönkkönen totesi, että selvityksen aikana ei arvioitu myöskään paljonko dataa on tutkijoiden käytössä esimerkiksi yrityspuolelta. Lisäksi lainsäädännölliset ja kustannuskysymykset rajattiin ulkopuolelle.

Puheenjohtaja painotti myös, että kokonaisuus ei rajoitu yliopistoihin. Korkeakoulujen sopimuskierroksella moni yliopisto nosti asian keskusteluun, joten selvitystä lähdettiin edistämään. Rahoituksen osalta todettiin, että kokonaisuus koskee kaikkia ministeriöitä ja että YTF on yksi KTPO:n CSC:ltä ostamista palveluista. Visioon muotoillaan myös tästä teemasta kirjaus ja keskustelua jatketaan taas seuraavassa kokouksessa.

## 5. Tutkimuksen digitalisaation tilannekuva ja skenaariotyö

Heikki Mannila esitteli tutkimuksen digitalisaation nykytilaa ja siihen liittyvää skenaariotyötä. Esityksessä korostettiin, että digitalisaatiolla on sekä suoria että epäsuoria vaikutuksia oppimiseen ja koulutukseen, ja näiden vaikutusten erottaminen toisistaan on usein haastavaa. Mannila nosti esiin tarpeen pohtia, mitä Suomessa kannattaa tehdä strategisesti digitalisaation edistämiseksi.

Tilannekuvasta pienryhmä totesi, että suomalainen tutkimus on menestynyt monin tavoin, mutta korkeakoulujen henkilöstörakenteessa on tapahtunut muutoksia, kuten tutkimuspuolen nimikkeiden karsintaa Aalto-yliopistossa yliopistolakiuudistuksen jälkeen. Esityksessä korostettiin monitieteisen osaamisen merkitystä ja nostettiin esiin kansallisten ja kansainvälisten RSE-yhdistysten rooli.

Mannila viittasi kahteen kansainväliseen raporttiin (Bidenin hallinnon Supercharging Research: Harnessing Artificial Intelligence ja Royal Societyn Science in the age of AI), joiden mukaan tutkimus muuttuu tekoälyn avulla helpommaksi, mutta nykykehityksen perusteella muutokset voivat olla myös radikaaleja. Uudet menetelmät muovaavat tutkimuskysymyksiä ja voivat parantavat luotettavuutta, mutta on jo nähtävissä myös trendi, että tekoälyn avulla saatetaan tuottaa runsain määrin esimerkiksi keskinkertaisia tutkimusjulkaisuja. Esityksessä korostettiin yleisten menetelmien ja alakohtaisen seurannan tärkeyttä. Työssä nostettiin esiin haasteita, kuten vertaisarvioinnin kuorma ja tieteen eriytyminen. Lisäksi esille tuotiin tutkijankoulutuksen ja laskentainfrastruktuurin ylläpidon merkitys, alojen välinen yhteistyö sekä opetusmenetelmien ja -sisältöjen kehittäminen.

Pienryhmän identifioimat skenaariot:

- Tutkimuksen digitalisaatio etenee ja kattaa aikaisempaa laajemman osan tutkimuksesta.
- Tekoälyn nopea kehitys jatkuu ja siitä kehittyy hyvin merkittävä tutkimuksen apuväline, joka helpottaa globaalien haasteiden ratkaisemista.
- Tekoälyn nopea kehitys jatkuu, mutta tutkimuksen kannalta tuloksena on suuri määrä huonotasoisista tutkimusta, joka ei vie tiedettä eteenpäin eikä ratkaise suuria ongelmia, vaan päinvastoin johtaa tutkimuksen tason laskemiseen.

- Geopoliittiset ja muut tekijät johtavat siihen, että edistyneimmät tekoälymenetelmät ovat vain tiettyjen maiden tai tiettyjen tutkijoiden käytettävissä; tiede eriytyy ja julkaiseminen polarisoituu.
- Tekoälyn nopea kehitys mullistaa kansainvälisen talouden ja merkittävä osa Euroopan asiantuntijatyöpaikoista katoaa tai siirtyy muualle. Tämän seurauksena länsimaat vajoavat syvään lamaan, levottomuudet leviävät ja tutkimuksen rahoittaminen loppuu.

Pienryhmä suosittaa, että tutkimuksen digitalisaation mukanaan tuomat muutokset tulee ottaa huomioon tutkimuksessa ja tutkijakoulutuksessa. Tutkimuksen digitalisaation edetessä korkeakoulut tarvitsevat entistä enemmän laskenta- ja tekoälymenetelmiä osaavaa henkilökuntaa, ja korkeakoulujen tulisikin pohtia urajärjestelmiään tästä näkökulmasta. Tutkimuksen digitalisaation ja tekoälymenetelmien tarvitsemasta laskentainfrastruktuurista tulee pitää jatkossakin huolta.

Lisäksi ryhmä suosittelee, että alojen välistä monitieteistä yhteistyötä tekoälymenetelmien kehityksessä ja soveltamisessa tulee kannustaa ja tukea johdonmukaisesti ja pitkäaikaisesti. Tekoälymenetelmien kehittyminen vaikuttaa tutkimuksen tekemisen lisäksi myös opetukseen ja opetusmenetelmiin monimutkaisilla tavoilla, jotka oletettavasti ovat eri aloilla erilaisia. Alojen välinen yhteistyö on tärkeää myös opetuksen näkökulmasta. Korkeakoulujen yhteistyö voisi olla hyödyllinen pohdittaessa esim. tekoälyn vaikutuksia jatkokoulutukseen ja myös alakohtaisia vaikutuksia selvittäessä. Tekoälymenetelmät vaikuttavat myös eri osaamisten kysyntään ja koulutustarpeisiin. Tekoälymenetelmien vastuullinen käyttö (sisältäen myös kestävyysnäkökulmat) tulee varmistaa. Tutkimuksen digitalisaation ja tekoälymenetelmien kehittyessä korkeakouluissa tarvitaan alakohtaisen kehittämisen ja strategisen lähestymistavan yhdistämistä.

Keskustelussa nousi esiin tekoälyn vaikutus koulutustarpeisiin, kuten matalan tason koodauksen ja kielenkääntämisen automatisoituminen. Mannila totesi, että suomalaisen tutkimuksen kannalta olisi strategisesti järkevää keskittyä alakohtaisiin tekoälyratkaisuihin, kuten physics-informed AI ja medicine-informed AI, sillä yleiskäyttöisten kielimallien kilpailussa Suomi ei pärjää suurille teknologiayrityksille. Pohdittiin myös kestävyysnäkökulmia, kuten sähkön saatavuutta ja hintaa, erityisesti suurten tekoälylaskentakeskusten osalta. Mannila totesi, että alakohtaista kehitystä on tärkeää seurata jatkuvasti, eikä digitalisaation kehityksessä ole vielä nähtävissä tasaantumista. Pienryhmän raportti on julkaistavissa hankeikkunassa oikolukukierroksen jälkeen.

Jukka Mönkkönen toi esiin mielenkiintoisen esityksen, jossa Chalmersin Ross King totesi, että Eurooppa on hävinnyt laajojen kielimallien kilpailun isoille teknologiayrityksille, mutta meille oiva näytön paikka olisi ”AI for science”. EU:ssa on osaamista ja etumatkaa mm. laboratoriorobotiikassa. Etulyöntiasemaa kannattaisi hyödyntää ja tämä alleviivaa myös datan ja sen hallinnan merkitystä. Mannila totesi myös, että suomalaisen tutkimuksen kannalta kiintoisaa olisi keskittyä esimerkiksi physics-informed tai medicine-informed tekoälyyn. Hän tarkensi, että alakohtaisella tarkastelulla raportissa tarkoitetaan tätä. Yleiskäyttöiset mallit eivät riitä, alakohtaisuutta tarvitaan ja esimerkiksi tärkeiksi voivat muodostua tiettyyn erityisalaan viritettyjen kielimallien käyttö yhdistettynä vaikkapa data-analyysin osaamiseen.

## **6. Mahdolliset muut asiat**

Laura jakaa Jukan ja Heikin esitykset ryhmälle, kunhan ne on viimeistelty.

## **7. Seuraava kokous**

Syksyn toinen kokous järjestetään 19.11. kello 10.30–12.00 opetus- ja kulttuuriministeriössä (Meritullinkatu 10). Siellä jatkuu korkeakoulutuksen ja tutkimuksen vision käsittely ryhmän teeman tiimoilta.

## **8. Kokouksen päättäminen**

Puheenjohtaja päätti kokouksen kello 14.05