CSC:n asiantuntijalausunto tulevaisuusvaliokunnalle (E 27/2024 vp)

Näkökulmia supertietokoneilla tehtävän tutkimuksen hyödyntämiseen EU:n terveysunionissa pitkällä aikavälillä

CSC – Tieteen tietotekniikan keskus Oy (CSC) kiittää mahdollisuudesta lausua suurteholaskennan eli supertietokoneilla tehtävän tutkimuksen hyödyntämisestä terveysunionissa. CSC pitää tervetulleina Suomen ennakkovaikuttamista Euroopan terveysunionin tulevaisuuteen sekä Suomen kantoja mm. keskittymisestä terveysalalla EU-sääntelyn laatuun ja sääntelyn ja muiden tavoitteiden toimeenpanoon, EU:n terveystietoavaruuden (EHDS) alaisten tietojen toisiokäytön mahdollistamiseen TKI-toiminnan hyväksi, aktiiviseen osallistumiseen miljoonan genomin hankkeeseen sekä ylipäänsä kokonaisvaltaiseen lähestymiseen terveysturvallisuuden, tautien hoidon ja ennaltaehkäisyn ja esimerkiksi lääkkeiden kehittämisen osalta.

Terveysdatan hyödyntäminen TKI-toiminnassa edellyttää mahdollistavaa lainsäädäntöä ja rakenteellisia uudistuksia

Suomen kilpailukyvylle on tärkeää, että suurteholaskentaa voidaan kattavasti hyödyntää tutkimuksessa terveysalan kasvattamiseksi. Kansallisen datan ja laskennan infrastruktuurin, EuroHPC-supertietokone LUMIn sekä sitä seuraavien suurteholaskennan investointien ansiosta Suomella onkin käytössään maailman kärkeä olevat TKI-työkalut terveysunionin ja esimerkiksi kansallisen terveys- ja hyvinvointialan kasvua ja uudistumista vauhdittavan TKI-ohjelman tavoitteiden edistämiseen. Tämän potentiaalin hyödyntämiseksi lainsäädännön ja määräysten on luotava mahdollisuuksia terveystiedon tietoturvalliselle käytölle terveysalan TKI-toiminnassa, joka on yhä kansainvälisempää ja yhä dataintensiivisempää.

Yhtäläisten osallistumismahdollisuuksien turvaamiseksi ja tutkimuksen edistämiseksi EHDS:ssä, sosiaali- ja terveystietojen toissijaista käyttöä säätelevää kansallista lakia on uudistettava siten, että aineistojen ja datan uudelleenkäyttö ja yhdistely helpottuvat. Toisiolain nojalla annettavaa määräystä tietoturvallisista käyttöympäristöistä tulee päivittää, jotta se sallii kansallisesti tehtyjen suurteholaskennan investointien käytön sensitiivisellä datalla, kuten terveysdatalla tehtävään tutkimukseen. Tämän rinnalla on jatkettava EU-tason vaikuttamista sen varmistamiseksi, että terveystietoavaruutta koskevan asetuksen nojalla asetettavat EU-tason vaatimukset turvallisille käyttöympäristöille toimivat samansuuntaisesti. Lisäksi on tärkeää varmistaa, että terveystiedon TKI-käytön näkökulma huomioidaan EU:n yleisen tietosuoja-asetuksen väliarviossa sekä hallitusohjelmassa linjatussa kansallisen tietosuojalainsäädännön uudistuksessa, sillä tietosuojasääntely asettaa olennaisia reunaehtoja terveystiedon hyödyntämiselle.

CSC – TIETEEN TIETOTEKNIIKAN KESKUS OY
Keilaranta 14, PL 405, 02101 Espoo, puh. 09 457 2001,
Y-tunnus 0920632-0, www.csc.fi

CSC – IT CENTER FOR SCIENCE LTD.
Keilaranta 14, P.O. BOX 405, FI-02101 Espoo, Finland, Tel. +358 9 457 2001
VAT number FI09206320, www.csc.fi

Terveysturvallisuudesta sairauksien parempaan ymmärtämiseen, diagnosointiin ja hoitoon – supertietokoneet unionin tavoitteiden mahdollistajina

Viimeistään COVID-19 -pandemia paljasti, että EU:n tavoitteet strategisesta autonomiasta sekä kansalaistensa hyvinvoinnin, terveyden ja turvallisuuden edistämisestä kytkeytyvät yhä enemmän biologisten ja lääke- ja terveystieteellisten alojen tutkimus-, kehitys- ja innovaatiotoimintaan sekä sitä mahdollistaviin data- ja tutkimusinfrastruktuureihin ja digitaalisiin teknologioihin. Esimerkiksi CSC:n suurteholaskentaresursseja on hyödynnetty viruksen ilmateitse tapahtuvan leviämisen mallintamiseen ja toimintamekanismin ymmärtämiseen suojautumistoimien parantamiseksi ja lääketutkimuksen pohjustamiseksi[1]. Eurooppalaiseen dataportaaliin[2] on kerätty ainakin kahdeksan miljoonaa koronavirussekvenssiä, minkä ansiosta infektion molekyylitason toiminnan ja virusvarianttien tutkimusta voidaan jatkaa tulevaisuuden varautumistarpeita varten.

Supertietokoneiden ympärille kehittyvät, esimerkiksi yhteiseurooppalaisten EuroHPC-supertietokonekeskusten kaltaiset ekosysteemit tuovat etua monitieteiselle tutkimukselle mahdollistaessaan eri alojen suurten data-aineistojen yhdistelyn ja käsittelemisen, mikä luo mahdollisuuksia myös E-kirjeessä korostetulle Yhteinen terveys (*one health)* -lähestymistavalle. Kajaanissa sijaitsevalla EuroHPC-supertietokone LUMIlla on jo käynnissä merkittäviä biodiversiteetti-, ympäristö- ja ilmastodataan perustuvia tutkimushankkeita[3] [4], ja yhdistelemällä tällaista tietoa terveys- ja lääketieteelliseen tietoon voidaan tulevaisuudessa oppia enemmän esimerkiksi lääkeresistenssin ja pandemioiden kaltaisista terveysuhista. Ilmastopuolella superkoneilla on jo tehty esimerkiksi sään ääri-ilmiöiden mallintamista[5] ja niitä voidaan tulevaisuudessa hyödyntää myös rajat ylittävien terveysuhkien, kuten vaikkapa säteilytilanteiden, yhä tarkempaan mallintamiseen.

Tarkempi sairauksien analyysi edellyttää monipuolisempaa dataa ja suurempaa laskentatehoa

Eri alojen suuria data-aineistoja yhdistelemällä on siis mahdollista saavuttaa merkittäviä tieteellisiä läpimurtoja esimerkiksi juuri terveystutkimuksessa. Dataintensiivinen tutkimus on yleistynyt jo jonkin aikaa, ja nykyään on tyypillistä myös perinteiseen rekisteritutkimukseen perustuvien havaintojen täydentäminen eri biologisella tiedolla, kuten proteiini-, genomi- tai muulla molekyylitason datalla. Tällaisella analyysillä voidaan yhä tarkemmin tunnistaa harvinaissairauksia ja sairauksien syntymekanismeja sekä kehittää yksilöllistetympiä hoitoja ja lääkityksiä, mikä kaikki tukee terveysunionin tavoitteita sairauksien ehkäisyn ja hoidon vahvistamisesta.

[1] https://csc.fi/uutinen/puhdilla-koronaa-vastaan-katsaus-covid-19-pikakaistan-tuloksiin/ ; https://www.aka.fi/en/about-us/whats-new/press-releases/20202/high-level-research-infrastructures-support-covid-19-research/
[2] https://www.covid19dataportal.org/the-european-covid-19-data-platform
[3] https://biodt.eu/
[4] https://stories.ecmwf.int/destination-earth/index.html
[5] https://stories.ecmwf.int/m-t-o-france-wins-bid-to-develop-destination-earth-s-on-demand-extremes-digital-twin/index.html

CSC – TIETEEN TIETOTEKNIIKAN KESKUS OY
Keilaranta 14, PL 405, 02101 Espoo, puh. 09 457 2001,
Y-tunnus 0920632-0, www.csc.fi

CSC – IT CENTER FOR SCIENCE LTD.
Keilaranta 14, P.O. BOX 405, FI-02101 Espoo, Finland, Tel. +358 9 457 2001
VAT number FI09206320, www.csc.fi

Terveystutkimuksessa käytettävien aineistojen monipuolistuminen ja monimutkaistuminen edellyttävät paitsi saatavilla olevan datan määrän ja laadun parantamista, myös sellaista tallennus- ja laskentakapasiteettia ja erityisosaamista, jota vain maailman tehokkaimmat supertietokonekeskukset, kuten LUMI ja sen ympärille rakentuva ekosysteemi, kykenevät tarjoamaan. Yhteiskunnallisesti vaikuttavan terveystutkimuksen edistämisen kannalta onkin hyvin positiivista, että Suomi tavoittelee terveysunionissa EHDS-toisiokäytön edistämistä sekä aktiivisuutta miljoonan genomin aloitteen toimeenpanossa ja on lisäksi jo nyt sitoutunut uuteen EuroHPC-supertietokonehankintaan nykyisen LUMIn tullessa käyttöikänsä päähän muutaman vuoden kuluttua.

Supertietokoneet ovat tekoälykehityksen kasvun edellytys myös tulevaisuudessa

Tarve suurteholaskennalle kasvaa koko ajan, kun tekoäly ja mutkikkaat matemaattiset mallit yleistyvät. Supertietokoneilla on erityisen keskeinen rooli tekoälyn kehittämisessä, joka perustuu viime kädessä suuriin ja laadukkaisiin data-aineistoihin. Uusien supertietokoneiden suuri laskentateho tekoälysovelluksissa nojaa GPU-prosessoreihin perinteisten CPU-prosessoreihin sijaan ja niillä on saatu lyhennettyä laskentaprosesseja alle kymmenykseen aiemmasta esim. kuvaperusteisten, syöpätutkimusta tukevien algoritmien kehityksessä. Tällä tavalla LUMI-supertietokone on edistänyt esimerkiksi syöpädiagnostiikan tehostamista [6] [7], mikä on lupaavaa EU:n syöväntorjuntasuunnitelmalle osana terveysunionia myös tulevaisuudessa tekoälytutkimuksen kehittyessä edelleen.

Tekoälyä voi hyödyntää hyvin laajasti terveysunionin tavoitteissa, ja esimerkiksi kielimalleihin perustuvan generatiivisen tekoälyn käyttö terveysalalla tulee hyödyttämään sekä hoito- että TKI-työtä. Sen avulla voidaan esimerkiksi nopeuttaa tiedonkeruuta hoitosuunnitelmien seurannassa tai tutkimuslähteiden tarkastamisessa.[8] Eurooppalaiselle terveysunionille on siten tärkeää, että kielimalleja kehitetään tulevaisuudessa yhä vahvemmin eurooppalaisilla kielillä alaa dominoivan englannin rinnalla. LUMIlla on jo kehitetty eurooppalaisia kielimalleja pohjatyöksi niiden laajentamista ja alakohtaista jatkojalostusta varten[9], mukaan lukien suomeksi[10] ja muilla pohjoismaisilla kielillä[11].

Yllä mainittu datalähteiden monipuolistuminen mm. EHDS:n ja miljoonan genomin hankkeen onnistuneen toimeenpanon seurauksena yhdistettynä suureen laskentatehoon luovat merkittävää lisäpotentiaalia, sillä generatiivisen tekoälyn kunnollinen hyödyntäminen hoito- ja diagnosointityön tukena vaatii yksityis- ja myös yksilökohtaisempaa dataa.[12] Samalla on huomattava, että inhimillinen

[6] https://www.lumi-supercomputer.eu/lumi-provides-new-opportunities-for-artificial-intelligence-research/
[7] https://www.hpcwire.com/2023/06/08/inside-compatai-one-of-lumis-first-projects/
[8] https://www.brookings.edu/articles/generative-ai-in-health-care-opportunities-challenges-and-policy/
[9] https://www.lumi-supercomputer.eu/now-language-models-will-be-trained-in-european-languages-on-one-of-the-worlds-largest-text-collections/
[10] https://www.lumi-supercomputer.eu/research-group-created-the-largest-finnish-language-model-ever-with-the-lumi-supercomputer/
[11] https://www.silo.ai/blog/viking-7b-13b-33b-sailing-the-nordic-seas-of-multilinguality
[12] Ks. viite 9.

CSC – TIETEEN TIETOTEKNIIKAN KESKUS OY
Keilaranta 14, PL 405, 02101 Espoo, puh. 09 457 2001,
Y-tunnus 0920632-0, www.csc.fi

CSC – IT CENTER FOR SCIENCE LTD.
Keilaranta 14, P.O. BOX 405, FI-02101 Espoo, Finland, Tel. +358 9 457 2001
VAT number FI09206320, www.csc.fi

osaaminen ja ymmärrys tekoälyn kehittämisestä ja hyödyntämisestä tutkimuksessa ja hoitotyössä vaativat panostuksia myös pelkän ”raudan”, datan ja teknologian lisäksi.

Panostukset tietoturvalliseen datanhallintaan ja mahdollistavaan sääntelyyn maksimoivat supertietokoneiden tuoman kilpailuedun terveysalan TKI-toiminnalle

Koska hyvin hallittu data on niin suurteholaskennan kuin tekoälyn kehittämisenkin raaka-ainetta, Suomen ja EU:n tulee panostaa datan saatavuuteen, uudelleenkäytettävyyteen ja hallintaan. *Lainsäädännön tulee tukea ja mahdollistaa näitä tavoitteita, jotta tutkimusinfrastruktuuri-investoinneista ja työkaluista, kuten supertietokoneista, saadaan kaikki hyöty irti terveysalan TKI-toiminnassa. Tulee varmistaa, että eurooppalaiset ja kansalliset vaatimukset ja määräykset datan käsittely-ympäristöistä, jollainen supertietokonekin on, tukevat sekä tietoturvan että käytettävyyden tavoitteita. Niiden tulee tulevaisuudessa mahdollistaa suurteholaskennan käyttö kattavasti ja monipuolisesti myös sellaisiin terveystietoihin, jotka ovat tietosuojasääntelyn ja toisiolain alaisia.* COVID -19 tuskin jää viimeiseksi pandemiaksi, ja EuroHPC-järjestelmät voivat olla keskeisessä roolissa sekä itse sairauksien tutkimisessa että terveydenhuollon resurssien optimointia koskevien laskelmien tekemisessä, mikäli ne saadaan osaksi tietoturvallisen datanhallinnan järjestelmää. Tämä tarkoittaa kansalaisista kerätyn terveystiedon laaja-alaista hyödyntämistä tavalla, jota ei välttämättä voida toteuttaa hyvinvointialueiden tai sairaaloiden omissa järjestelmissä. Myös tekoälyn vaikuttava hyödyntäminen hoitotyössä suomen kielellä tulee edellyttämään suurteholaskennan sallimista nykyistä monipuolisemmille datalähteille.

Supertietokoneet ovat keskitettyjä mutta monikäyttöisiä, jaettuja järjestelmiä, minkä takia vaaditaan mekanismeja suojaamaan arkaluonteista henkilötietoa muilta käyttäjiltä ja ulkopuolisilta. *Moni uusi tietoturvamekanismi on vasta kehitteillä, joten määräysten tulee asettaa ja kuvailla tietoturvan tila ja tavoitteet määräämättä kuitenkaan yksityiskohtaisesti keinoja, joilla ne tulee saavuttaa. Esimerkiksi nykyisten, toisiolain nojalla annettavien määräysten kategorinen kielto internetyhteyksistä henkilötietoon perustuvan terveystiedon käsittelyssä sulkee pois monta potentiaalista salausmenetelmää sekä käytännössä nykyisten supertietokoneresurssien tehon hyödyntämisen toisiolain alaisessa tutkimuksessa.*

Teknologianeutraalius on ollut EU-sääntelyssä vallitseva periaate, ja tätä ajatusta tulee noudattaa myös kansallisella tasolla. Periaatteellisella tasolla on pohdittava, sopiiko salaisten aineistojen suojaamiseen tehty Katakri-työkalu täysin lähtökohdaksi tieteelliseen tutkimuskäyttöön liittyvien järjestelmien arviointiin. *Hallitusohjelman mukaiset tietosuojalainsäädännön ja toisiolain uudistukset, EHDS sekä käsittely-ympäristöjen vaatimusten päivitys tarjoavat mahdollisuuden ottaa tietoturvalliseen käyttöön suurteholaskennan tutkimusresurssit, joiden merkitys suomalaisen ja eurooppalaisen terveysalan tutkimukselle ja tulevaisuudelle on kiistaton.*

Lainsäädännön tarkastelun ja esteiden purkamisen lisäksi on tärkeää kehittää entistä toimivampia yhteyksiä eurooppalaisten datainfrastruktuurien ja EuroHPC-supertietokonekeskusten välillä sekä hajautetun arkkitehtuurin avulla että tavoittelemalla datainfrastruktuurien sijoittautumista

supertietokoneiden yhteyteen sekä yleisesti että bio- ja terveystieteissä[13]. Tietoturvallisen datanhallinnan ja laskennan kehitystarpeet tutkimuskäyttöä varten tulevat painottumaan EHDS:n myötä, ja tällainen työ on Suomessa jo käynnissä[14]. Suomi on jo nyt edelläkävijä datanhallintainfrastruktuurien ja supertietokoneiden yhdistämisessä kansallisen datanhallinnan ja laskennan yhteishankkeen myötä sekä operoidessaan bioinformatiikkaan erikoistunutta Suomen ELIXIR-osakeskusta[15]. Yleisesti ottaen suurteholaskennan käyttö tietosuojan tai aineettomien oikeuksien alaisille aineistoille esimerkiksi ennustavien, ml. tekoälymallien tapauksessa edellyttää, että palvelukehityksessä on ison laskentakapasiteetin lisäksi käyttötarkoitusta huomioivaa datanhallinnan ja tietosuojan osaamista.

Espoossa, 21.5.2024

CSC-Tieteen tietotekniikan keskus Oy

Kimmo Koski

Toimitusjohtaja

Irina Kupiainen

Yhteiskuntasuhdejohtaja

[13] Vrt. https://csc.fi/uutinen/cscn-raskaan-laskennan-ymparisto-laajenee-massiivisella-datanhallintajarjestelmalla/

[14] https://csc.fi/blogi/parempia-yksilollisia-hoitomenetelmia-kehittyneiden-tiedonsiirto-ja-analyysimenetelmien-avulla/

[15] https://elixir-europe.org/; Valtiosopimus 7/2015