

Federoidun laskennan mahdollistaminen sosiaali- ja terveystietojen toissijaisessa käytössä

Taustapaperi 7.2.1

FinHITS — Strengthening Finnish Health Data ICT for Secondary Use

Projektinumero 101126512

21. tammikuuta 2025



**Euroopan unionin
osarahoittama**

FINDATA

Euroopan unionin osarahoittama. Esitetyt näkemykset ja mielipiteet ovat ainoastaan tämän tekstin laatijoiden näkemyksiä eivätkä välttämättä vastaa Euroopan unionin tai Euroopan terveys- ja digitaaliasioiden toimeenpanoviraston (HaDEA) kantaa. Euroopan unioni ja HaDEA eivät ole vastuussa niistä.

Sisällys

1	Tiivistelmä.....	3
2	Johdanto.....	3
2.1	Tausta.....	3
2.2	Tavoitteet.....	3
2.3	Menetelmät.....	4
3	Nykyinen lainsäädäntö.....	5
4	Federoidun laskennan määritelmät.....	6
5	Keskeiset käyttötapaukset.....	7
5.1	Tietojohtamisen kansainvälinen verkosto.....	7
5.1.1	Kuvaus käyttötapauksesta.....	7
5.1.2	Käyttötapauksen merkitys.....	7
5.1.3	Toteutusmahdollisuudet nykyisessä toimintaympäristössä.....	7
5.1.4	Jatkokehitystarpeet.....	8
5.2	Federoitu oppiminen eurooppalaisessa yhteistyöhankkeessa.....	8
5.2.1	Kuvaus käyttötapauksesta.....	8
5.2.2	Käyttötapauksen merkitys.....	8
5.2.3	Toteutusmahdollisuudet nykyisessä toimintaympäristössä.....	8
5.2.4	Jatkokehitystarpeet.....	9
5.3	Hybridi federoidun oppimisen tutkimus (horisontaalinen ja vertikaalinen).....	9
5.3.1	Kuvaus käyttötapauksesta.....	9
5.3.2	Käyttötapauksen merkitys.....	9
5.3.3	Toteutusmahdollisuudet nykyisessä toimintaympäristössä.....	9
5.3.4	Jatkokehitystarpeet.....	10
6	Yhteenveto keskeisistä haasteista ja suositukset jatkokehitykseen.....	10



Asiakirjan tiedot

Kirjoittajat	Organisaatio
Heikki Lanu, tietoturvapäällikkö	Findata – Sosiaali- ja terveysalan tietolupaviranomainen
Maari Parkkinen, kehittämisspäällikkö	Findata – Sosiaali- ja terveysalan tietolupaviranomainen
Hanna Tervonen, data manageri	Findata – Sosiaali- ja terveysalan tietolupaviranomainen
Hannu Vilpponen, ICT päällikkö	Findata – Sosiaali- ja terveysalan tietolupaviranomainen
Mika Hilvo, tutkimustiimin vetäjä	Teknologian tutkimuskeskus VTT Oy
Miikka Kallberg, kehityspäällikkö	CSC – Tieteen tietotekniikan keskus Oy
Jaakko Lähtenmäki, johtava tutkija	Teknologian tutkimuskeskus VTT Oy
Juha Pajula, erikoitutkija	Teknologian tutkimuskeskus VTT Oy
Emmi Turunen, tuotepäällikkö	HUS
Arho Virkki, analytiikkajohtaja	Varsinais-Suomen hyvinvointialue

Asiantuntijat	Organisaatio
Miikka Kallberg, kehityspäällikkö	CSC – Tieteen tietotekniikan keskus Oy
Heikki Lanu, tietoturvapäällikkö	Findata – Sosiaali- ja terveysalan tietolupaviranomainen
Maari Parkkinen, kehittämisspäällikkö	Findata – Sosiaali- ja terveysalan tietolupaviranomainen
Hanna Tervonen, data manageri	Findata – Sosiaali- ja terveysalan tietolupaviranomainen
Hannu Vilpponen, ICT päällikkö	Findata – Sosiaali- ja terveysalan tietolupaviranomainen
Annu Kaila	HUS
Emmi Turunen, tuotepäällikkö	HUS
Harri Rantala, datatieteilijä	Pirkanmaan hyvinvointialue
Tapio Pahikkala, professori	Turun yliopisto
Annika Pirnes, palvelupäällikkö	Varsinais-Suomen hyvinvointialue



Sosiaali- ja terveysalan tietolupaviranomainen

21. tammikuuta 2026

Per-Erik Gustafsson, IT-suunnittelija	Varsinais-Suomen hyvinvointialue
Arho Virkki, analytiikkajohtaja	Varsinais-Suomen hyvinvointialue
Antti Yli-Karhu, IT-suunnittelija	Varsinais-Suomen hyvinvointialue
Mika Hilvo, tutkimustiimin vetäjä	Teknologian tutkimuskeskus VTT Oy
Jaakko Lähteenmäki, johtava tutkija	Teknologian tutkimuskeskus VTT Oy
Juha Pajula, erikoistutkija	Teknologian tutkimuskeskus VTT Oy

Muutoshistoria

Päiväys	Versio	Muokkaaja	Status
26.3.2025	0.1	Maari Parkkinen	Luonnos
4.4.2025	0.1	Hanna Tervonen	Luonnos
3.6.2025	0.2	Maari Parkkinen	Luonnos
27.6.2025	0.2	Hanna Tervonen	Luonnos
10.7.2025	0.2	Maari Parkkinen	Luonnos
27.10.2025	0.2	Hanna Tervonen	Luonnos
4.11.2025	0.2	Maari Parkkinen	Luonnos
21.1.2026	0.3	Fiona Helminen	Valmis



1 Tiivistelmä

Taustapaperi tarkastelee federoidun laskennan mahdollisuuksia sosiaali- ja terveystietojen toissijaisessa käytössä erityisesti tietoturvalisten käyttöympäristöjen kontekstissa. Federoitu laskenta mahdollistaa henkilötietojen käsittelyn ilman aineistojen siirtoa, mikä tukee tietosuojaa ja rekisteröityjen oikeuksia. Raportti esittelee federoidun laskennan määritelmät, keskeiset käyttötapaukset sekä lainsäädännölliset reunaehdot Suomessa ja EU:ssa. Lisäksi se tarjoaa asiantuntijaryhmän suosituksia jatkokehitykselle, korostaen datan harmonisoinnin, teknisten ratkaisujen ja sääntelyn selkeyttämisen merkitystä federoidun laskennan tehokkaaksi hyödyntämiseksi kansallisessa ja kansainvälisessä tutkimuksessa. Raportin laadintaan on osallistunut laaja asiantuntijajoukko, jolla on käytännön kokemusta federoidun laskennan haasteista ja mahdollisuuksista eri toimintaympäristöissä.

2 Johdanto

2.1 Tausta

Federoitu laskenta mahdollistaa henkilötietojen käsittelyn ilman, että dataa on tarpeen koostaa ja siirtää yhteen käsittely-ympäristöön.

Federoitu laskenta voi olla erityisesti tarpeen, kun käsitellään

- 1) erittäin suuria määriä tietoja, joiden siirtäminen on aineiston koon takia haastavaa,
- 2) erityisiin henkilötietoryhmiin kuuluvia tietoja, joiden käsittely-ympäristö voi olla hyvin rajattu tai
- 3) kun eri maiden tietoja halutaan hyödyntää, mutta vaihtelevat lainsäädännöt ja käytänteet rajoittavat omien kansalaisten tietojen siirtoa toiseen maahan.

Federoidussa laskennassa on etuja. Esimerkiksi tietosuojan näkökulmasta on hyvä, että aineistoja voidaan käsitellä ilman niiden koostamista yhdelle alustalle. Henkilötietojen rajatumpi käsittely tukee myös rekisteröityjen oikeuksia.

Federoidun laskennan toteuttamisen edellytys on yhtenäisten tietomallien käyttö.¹ Analyseissä käytettävä aineisto tulee harmonisoida ja tallentaa samanlaiseen tietomalliin, jolloin paikallisesti sijaitsevien aineistojen analysointi onnistuu yhteisellä, jaetulla algoritmilla. Potilastietojen harmonisoinnissa yleisesti käytetty tietomalli on OMOP (Observational Medical Outcomes Partnership). Suomalaisessa FinOMOP -hankkeessa on tuotettu OMOP-pohjaisia yhtenäisiä käytäntöjä, joita sovelletaan laajasti ympäri maan.

Tämä taustapaperi on kirjoitettu osana FinHITS – Strengthening Finnish Health Data ICT for Secondary use -projektin (jatkossa FinHITS-projekti) työpakettia 7, Tietoturvalinen käyttöympäristö (SPE). FinHITS-projekti on saanut rahoitusta EU4Health-ohjelmasta (Direct grants to Member States: for setting up services by Health Data Access Bodies -Secondary use of health data). Yksi työpaketin tavoitteista on edistää federoidun laskennan toteutusmahdollisuuksia tietoturvalisissa käyttöympäristöissä.

2.2 Tavoitteet

Tämän taustapaperin tavoitteena on kuvata, miten federoitua laskentaa voidaan toteuttaa kansallisesti ja kansainvälisesti sosiaali- ja terveystietojen toissijaisessa käytössä. Tarkastelua

¹ Laitinen T, Virkki A, Porkka K: FinOMOP: terveystietojen kansainvälinen harmonisointi. Duodecim 2022;138:1761–3.



tehdään erityisesti tietoturvalisten käyttöympäristöjen näkökulmasta. Tavoitteena on myös edistää federoidun laskennan tehokasta hyödyntämistä ja tukea toisioalaisia määriteltyjen auditoitujen tietoturvalisten käyttöympäristöjen yhteistä kehitystyötä.

2.3 Menetelmät

Taustapaperi on kirjoitettu kansallisten asiantuntijoiden yhteistyönä. Taustapaperissa esitetään yhtenäiset määritelmät federoidun laskennan eri toteutustavoille ja kuvataan keskeisten käyttötapauksen kautta kansallisessa ympäristössä esiintyviä haasteita federoidun laskennan toteuttamiselle. Käyttötapauksen pohjalta on lisäksi listattu keskeisiä jatkokehitystarpeita sekä mahdollisia ratkaisuja ja annettu suosituksia jatkokehityksen näkökulmasta.

Taustapaperin työstäminen on edennyt useammassa vaiheessa. Findata järjesti 6.6.2024 taustoittavan tapaamisen yhdessä Varsinais-Suomen hyvinvointialueen, Helsingin ja Uudenmaan sairaanhoitopiirin (HUS) ja Suomen molekyyli lääketieteen instituutin (FIMM) kanssa. Kyseiset organisaatiot olivat mukana pilottitutkimuksessa, jossa testattiin federoidun mallin mahdollisuuksia toisiokäytössä.² Tapaamisessa nostettuja keskeisiä näkökulmia olivat:

- Tarve yhtenäisille toimintatavoille federoidun laskennan tehokkaaksi toteuttamiseksi.
- Prosessien jouhevoittaminen ja nopeuttaminen.
- Tietoturvalisten käyttöympäristöjen optimointi tukemaan paremmin federoituja laskentaa, painottaen erityisesti anonymien parametrien vaihtoa eri ympäristöjen välillä ensisijaisena kehitystarpeena.

Taustapaperissa esitetyt käyttötapaukset sekä niiden perusteella muodostetut suositukset koostettiin keskeisten kansallisten asiantuntijoiden yhteistyönä kahdessa erillisessä työpajassa, joihin osallistui 19 henkilöä. Taustapaperin työstöön osallistuneet henkilöt edustivat yhteensä seitsemää organisaatiota (CSC - Tieteen tietotekniikan keskus Oy, Findata – Sosiaali- ja terveysalan tietolupaviranomainen, HUS-yhtymä, Pirkanmaan hyvinvointialue, Turun yliopisto, Varsinais-Suomen hyvinvointialue ja Teknologian tutkimuskeskus VTT Oy). Työpajojen aiheet ja ajankohdat olivat seuraavat:

- Työpaja 1: Nykytilanne ja keskeiset käyttötapaukset (to 20.3. klo 10–14)
- Työpaja 2: Jatkokehitystarpeet ja ratkaisuehdotukset (to 14.4. klo 9–11).

Työpajojen tavoitteena oli:

- tunnistaa keskeiset federoidun laskennan käyttötapaukset
- kuvata toteutus ja siihen liittyvät haasteet nykyisessä toimintaympäristössä
- tunnistaa jatkokehitystarpeet.

Taustapaperin sisällöt koostettiin työpajatyöskentelyn pohjalta ja taustapaperi lähetettiin yhteensä kolme kertaa osallistujille kommentoitavaksi ja täydennettäväksi työpajojen välissä sekä niiden jälkeen. Taustapaperiin pyydettiin myös kommentit sosiaali- ja terveysministeriön asettamalta korkean tason asiantuntijaryhmältä, jonka tehtävänä on laatia anonymisointia, tietosuoja ja tietoturva koskevat Tietolupaviranomaisen toiminnan periaatelinjaukset.

² Sitra: Lisänäyttöä uusista lääkkehoidoista -pilottiprojektin loppuraportti, 2024. Saatavilla: <https://www.sitra.fi/wp/wp-content/uploads/2024/04/lisanaytto-uusista-laakehoidoista.pdf>

3 Nykyinen lainsäädäntö

Suomessa sosiaali- ja terveystietojen hyödyntämistä toissijaisiin käyttötarkoituksiin säätelee laki sosiaali- ja terveystietojen toissijaisesta käytöstä 552/2019 (jatkossa toisiolaki). Toisiolaki säätelee lain alaisten tietojen käyttötarkoituksista, luvitusprosessista sekä käsittelystä.

Kirjoitushetkellä käynnissä on toisiolain korjaushanke, jossa esitetyt muutosehdotukset liittyvät tietolupien, tietopyyntöjen ja tietoaineistojen käsittelyyn, kansainväliseen tutkimusyhteistyöhön, toisiokäytön maksuihin, tietoaineistojen ja tulosten anonymisointiin sekä toisiolain suhteeseen klinisiin tutkimuksiin. Toisiolain ohella Suomessa on myös muuta sosiaali- ja terveystietojen toisiokäytön mahdollistavaa lainsäädäntöä. Tämä taustapaperi perustuu syksyllä 2025 voimassa olevaan toisiolakiin. Taustapaperia voidaan päivittää tarvittaessa.

Toisiolain perusteella yksilötasoisia aineistoja voidaan hyödyntää tieteellisen tutkimuksen, tilastoinnin, viranomaisten suunnittelu- ja selvitystehtävien, opetuksen sekä tietojohdamisen käyttötarkoituksissa. Tietojohdamista ja sosiaali- ja terveydenhuollon viranomaisohjausta voidaan tietyissä tilanteissa tehdä ilman tietolupaa (toisiolaki 41 §, 42 §). Tietojohdamisen käyttötarkoitukseen aineistoa ei tarvitse käsitellä toisiolain mukaisessa tietoturvalisessa käyttöympäristössä.

Jos tarkoituksena on yhdistellä useampia 6 §:n julkisen sosiaali- ja terveysalan rekisterinpitäjän tietoja, yhden tai usean yksityisen sosiaali- tai terveydenhuollon palvelunjärjestäjän rekisteritietoja tai Kanta-palveluihin tallennettuja asiakastietoja, myöntää tietoluvan Findata (44 §). Muussa tapauksessa luvan myöntää viranomainen, jonka tietoja haetaan.

Tietolupa perustuu yksilötasoinen aineisto luovutetaan toisiolain mukaisesti luvansaajan käsiteltäväksi aina toisiolain 20 §:ssä tarkoitettuun tietoturvaliseen käyttöympäristöön (51 §). Jos yksilötasoisien tietojen hyödyntäminen ei edellytä tietolupaa, ei tietojen hyödyntäjällä ole velvoitetta käyttää tietoturvalista käyttöympäristöä.

Tietoturvaliset käyttöympäristöt ovat tietoturvalisuuden arviointilaitosten auditoimia ympäristöjä, jotka täyttävät toisiolain ja Findatan antaman määräyksen (1/2022) vaatimukset. Määräyksen mukaisesti käyttöympäristön valvonnassa tulee kiinnittää erityistä huomiota tietoturvan valvontaan. Käyttöympäristön tietoturvan ajantasaisuutta tulee valvoa ja katselmoida säännöllisesti. Palveluntarjoajan tulee myös varmistaa vastaanottamiensa aineistojen eheys ja virheettömyys sekä tietoturvalisuus. Tietoluvan nojalla luovutettujen tietojen pohjalta julkaistavien tulosten anonymiteetin varmistaa Findata (52 §).

Tietoturvalisen käyttöympäristön vaatimukset eivät estä federoidun laskennan toteutusta edellyttäen, että tiedonhyödyntämisuunnitelma kattaa kyseessä olevan käyttötapauksen, käyttöympäristön ulkopuolelle ei siirry henkilötietoja ja arviointilaitos on hyväksynyt ratkaisun toteutuksen. Käyttöympäristön palveluntarjoajan vastuulla on varmistaa, ettei henkilötietoaineistoja lähde ulos käyttöympäristöstä. Tietoliikenneyhteys käyttöympäristöstä tai yksittäisestä käyttäjäympäristöstä voidaan avata, jos seuraavat asiat on huomioitu³:

- Yhteyden tarve on perusteltu ja yksittäisen käyttäjäympäristön tapauksessa yhteyden käyttötarkoitus ei ole ristiriidassa tietoluvan ehtojen kanssa.
- Yhteyden kohdeosoitteen osapuoli on tunnistettu ja arvioitu luotettavaksi.
- Yhteyden perustaminen, hallinta ja valvonta tapahtuu palveluntarjoajan toimesta.

³ Sitra: Lisänäyttöä uusista lääkkehoidoista -pilottiprojektin loppuraportti, 2024. Saatavilla: <https://www.sitra.fi/wp/wp-content/uploads/2024/04/lisanaytto-uusista-laakehoidoista.pdf>



Sosiaali- ja terveysalan tietolupaviranomainen

21. tammikuuta 2026

- Yhteys on suojattu palomuurilla ja vain käyttötapauksen edellyttämä verkkoliikenne sallitaan.
- Yhteydellä ei saa siirtyä henkilötietoja.
- Olennaisista muutoksista käyttöympäristöön on tehtävä ilmoitus arviointilaitokselle.

Käyttöympäristöjen toimintaa määrittää toisilain lisäksi kyberlaki. Kyberlaki asettaa Findatan määräystä tarkempia vaatimuksia mm. tietoturvan valvontaan. Erityisesti rekisteröityjen oikeuksien toteutumisen näkökulmasta on tärkeää huomioida myös EU:n yleinen tietosuoja-asetus (2016/679).

Asetus eurooppalaisesta terveystietoalueesta (myöh. EHDS-asetus) tuli voimaan 26.3.2025. EHDS:n tavoitteena on mahdollistaa rajat ylittävä terveystietojen toisiokäyttö yhtenäisillä hallintokäytännöillä, prosesseilla ja infrastruktuurilla. EHDS-asetus kattaa sekä ensiö- että toisiokäytön ja toisiokäytön osalta sitä sovelletaan suurimpaan osaan tietoluokista 26.3.2029 alkaen.

Myös EHDS-asetuksen mukaisessa toisiokäytössä on hyödynnettävä tietoturvalista käyttöympäristöä (Artikla 73). EHDS:n mukaisten tietoturvalisten käyttöympäristöjen vaatimukset määritetään 26.3.2027 mennessä annettavalla täytäntöönpanosäädöksellä. Parhailaan Sitran koordinoimassa kansainvälisessä TEHDAS2-hankkeessa luodaan suosituksia mm. tietoturvalisen käyttöympäristön teknisistä vaatimuksista tarkoituksena tukea täytäntöönpanosäädöksiä laadintaa.

Toisiokäyttöä koskevan EHDS-asetuksen luvussa IV on kuvattu prosessit, joiden mukaisesti yksilötasoiset tiedot kerätään ja siirretään tietoturvaliseen käyttöympäristöön. On kuitenkin odotettavissa, että jäsenmaat tulevat suhtautumaan vaihtelevin tavoin rajat ylittäviin tietojen siirtoihin, jossa tarkoituksena on kerätä tietoja yhteen tietoturvaliseen käyttöympäristöön. EHDS-asetuksen soveltamisen myötä tarve federoidun laskennan käytölle säilyy tai jopa lisääntyy, kun aineistojen saatavuus paranee. Federoitua laskentaa ei ole kuitenkaan suoraan huomioitu itse asetuksessa.

Parhailaan Euroopassa on käynnissä useita kansainvälisiä projekteja, joissa pyritään ratkaisemaan federoidun laskennan kautta erityyppisten arkaluonteisten aineistojen tutkimuskäyttöön liittyviä haasteita. Suurin osa näistä projekteista on saanut EU-rahoitusta. Ne keskittyvät yleensä yhden tai muutaman käyttötapauksen ratkaisemiseen, mutta eivät välttämättä vuorovaikuta keskenään tai huomioi tulevan EHDS-asetuksen vaatimuksia.

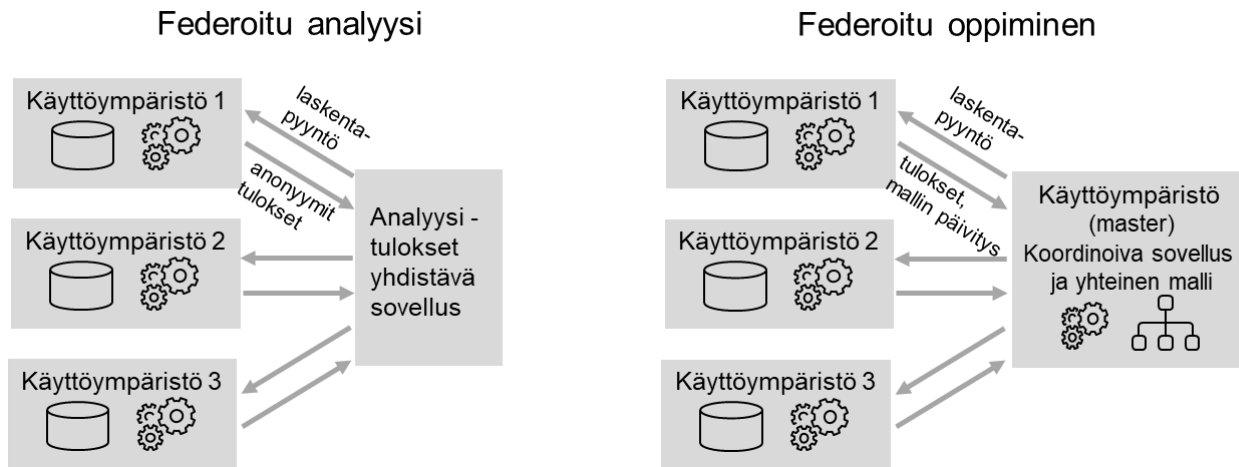
4 Federoidun laskennan määritelmät

Federoitu laskenta tarkoittaa laskentaa, jossa aineistoja käsitellään paikallisesti ilman, että kaikki aineistot kootaan yhteen käsittely-ympäristöön⁴. Federoiduksi laskennaksi voidaan katsoa niin kertaluontoiset laskennat kuin iteratiivinen tietojen käsittely (Kuva 1). Tässä taustapaperissa federoitu laskenta määritellään seuraavasti:

- **Federoitu laskenta** (eng. federated processing tai computing): yleistermi federoidulle laskennalle
- **Federoitu analyysi** (eng. federated analysis): kertaluontoinen federoitu laskenta
- **Federoitu oppiminen** (eng. federated learning): iteratiivinen federoitu laskenta.

⁴ Federoidun laskennan rinnalla käytetään joskus termiä hajautettu laskenta. Hajautettu laskenta viittaa yleisesti menetelmiin, jotka tehostavat laskentaresurssin käyttöä, esimerkiksi rinnakkaislaskentaan tai High Performance Computing -menetelmiin. Tämän vuoksi tässä dokumentissa ei käytetä termiä hajautettu laskenta.





Kuva 1. Havainnollistava kuva federoidun analyysin ja federoidun oppimisen eroista.

5 Keskeiset käyttötapaukset

Työpajoissa nostettiin esille seuraavat keskeiset käyttötapaukset, joiden toteutuksen mahdollisuuksia käydään yksityiskohtaisesti läpi tässä yhteenvedossa. Käyttötapaukset edustavat esimerkkejä yleisimmistä käyttötapauksista, joiden esittelyn tarkoitus on kartoittaa yksityiskohtaisemmin toteutuksen mahdollisuuksia ja haasteita. Lista ei ole kattava ja muitakin merkittäviä käyttötapauksia on olemassa.

5.1 Tietojohdamisen kansainvälinen verkosto

5.1.1 Kuvaus käyttötapauksesta

Tavoitteena on luoda kansainvälinen tietojohdamisen verkosto, jossa voidaan suorittaa jatkuvaluonteista vertaisarviointia ja seuranta. Tietojohdamisen ja analytiikan avulla mahdollistetaan eri toimijoiden yhteistyö kansainvälisessä toimintaympäristössä. Seuranta voidaan toteuttaa automaattisesti osana jatkuvaa prosessia. Toimijoina voivat olla sote-palveluntuottajat, jotka käyttävät rakenteellisesti yhtenevää dataa.

5.1.2 Käyttötapauksen merkitys

Tietojohdamisen kansainvälinen verkosto mahdollistaisi pienten, otoskooltaan rajallisten selvitysten toteuttamisen yhdistämällä useiden toimijoiden dataa. Esimerkiksi monet vaativat toimenpiteet on kansallisesti keskitetty monessa maassa yhteen sairaalaan. Kun hoitoa halutaan kehittää, yhteistyökumppanit ovat ulkomaisia. Käyttötapauksista voidaan hyödyntää myös esimerkiksi pandemiatilanteessa ja organisaatioiden ohjauksessa.

5.1.3 Toteutusmahdollisuudet nykyisessä toimintaympäristössä

Toteutus on toisilain näkökulmasta mahdollista nykyisessä toimintaympäristössä, sillä tietojohdamista voi tehdä tietoturvallisten käyttöympäristöjen ulkopuolella. Lisäksi

Sosiaali- ja terveysalan tietolupaviranomainen

21. tammikuuta 2026

tietojohdantamiseen ei tarvita erillistä tietolupaa, kun sosiaali- ja terveydenhuollon palvelunantaja käyttää sen omassa toiminnassa syntyneitä tai sen omiin rekistereihin tallennettuja asiakastietoja, jos käyttö on välttämätöntä palvelunantajan vastuulla toteutettavan palvelutoiminnan tuottamista, seuranta, arviointia, suunnittelua, kehittämistä, johtamista ja valvontaa varten (toisilaki 41 §). Tällöin ei kuitenkaan ole mahdollista yhdistää yksilötasoisia tietoja, joten datan rakenteiden ja tietomallien harmonisointi on keskeinen haaste. Toisilaki ei ota kantaa tietojohdantamisen tuloksena syntyneiden tulosten anonymiteettiin.

HUS:ssa on kehitteillä toteutus federoidusta verkostosta, joka mahdollistaa teknisesti yllä kuvatun kaltaisen käyttötapauksen. Ratkaisu perustuu avoimeen lähdekoodiin (federoitu noodi, joka vastaa analyysien ajosta ja tulosten palauttamisesta) sekä GitHub Enterprise -palveluun, jonka toiminnallisuuksia hyödynnetään analyysikoodin ja tulosten toimittamiseen liittyvässä logistiikassa. Tässä käyttötapauksessa tietojohdantamista ei tehdä täysin hajautetusti, vaan se perustuu osin keskitettyihin palveluihin. Arkkitehtuuriratkaisuja tehtäessä todettiin, että verkoston skaalautuvuuden kannalta on kriittistä, että järjestelmä rakentuu näin.⁵

5.1.4 Jatkokehitystarpeet

Tämän käyttötapauksen toteuttaminen edellyttää datan harmonisointia, tehokkaita tietojohdantamisen käytäntöjä sekä lainsäädännön, tietosuojan ja tietoturvan huomioimista.

5.2 Federoitu oppiminen eurooppalaisessa yhteistyöhankkeessa

5.2.1 Kuvaus käyttötapauksesta

Tavoitteena on tehdä federoitua oppimista tutkimuksessa, jossa hyödynnetään aineistoja useammasta EU-maasta ilman datojen siirtämistä yhteiseen käsittely-ympäristöön. Tutkimusryhmä hakee tutkimukseen tietoluvan jokaisesta maasta erikseen. Kategoriset ja numeeriset muuttujat harmonisoidaan OMOP-muotoon. Signaali- ja kuva-aineisto on yhteismitallisessa EDF ja DICOM-muodossa.

Tutkimuksen tavoitteena voi olla esimerkiksi kehittää eurooppalainen keuhkosityöpäennustemalli tai toteuttaa kardiologian diagnostiikan tueksi algoritmi, joka pisteyttää mahdollisimman tehokkaasti erilaisia hoitomuotoja ja ennustaa hoitovastetta. Laskentamenetelmää voitaisiin käyttää tulevaisuudessa lääkärin diagnostiikan tukena. Koska laskenta perustuu syviin neuroverkkoihin, sitä ei voi validoida manuaalisesti.

5.2.2 Käyttötapauksen merkitys

Käyttötapaus kuvaa hyvin yleistyvissä määrin tapahtuvaa kansainvälisessä yhteistyössä tehtävää tutkimusta. Parempi hoitovasteiden ennustaminen auttaa potilasta ja vähentää terveydenhuollon kustannuksia. Kun opetusaineistoa on paljon, diagnostinen algoritmi voidaan opettaa herkäksi (sensitiivinen) ja tarkaksi (spesifi). Tällaista tutkimusta kannattaa tehdä kansainvälisenä yhteistyönä, koska yksittäisen maan dataa on liian vähän. Jos eri maiden aineistoja ei voi siirtää rajojen ulkopuolelle yhteiseen käsittely-ympäristöön, federoitu laskenta mahdollistaa tutkimuksen toteutuksen.

5.2.3 Toteutusmahdollisuudet nykyisessä toimintaympäristössä

Käyttötapaus ei tällä hetkellä ole toteutettavissa, sillä se vaatii yhteyden tietoturvalisesta käyttöympäristöstä data-alueeseen ja automaattisen, iteratiivisen parametrien/gradienttien siirron.

⁵ Lisää tietoa aiheesta: <https://phems.eu>



Vaihtoehtona on, että tutkimusryhmä sovittaa algoritmin jokaisessa maassa erikseen eli toteuttaa federoidun analyysin kertosovituksena ja yhdistää mallit jälkikäteen.

Käyttötapauksessa ei ole tarkoitus siirtää henkilötietoja ja siltä osin hanke on nykyisen toisiolain puolesta mahdollinen toteuttaa. Ympäristöjen tulisi olla rakennettu niin, että ne tukevat federoitua oppimista. On kuitenkin todennäköistä, että vaikka tekninen ratkaisu onnistuisi, esiintyisi haasteita tulosten anonymiteetin varmistuksen osalta. Algoritmit olisi voitava tarkistaa, jotta voidaan varmistua käyttöympäristöjen välillä liikkuvien tietojen anonymiteetista.

Hankkeessa hyödynnetään harmonisoitua OMOP-muotoista dataa. Paikalliset jakaumat ovat kuitenkin todennäköisesti hyvin erilaisia maiden välillä johtuen eroista esimerkiksi tupakoinnissa tai tietojen kirjauskäytännöissä. Signaali- ja kuva-aineisto on yhteismitallistettu EDF ja DICOM-muotoon. DICOM-standardista on kuitenkin monta eri versiota, ja tästä syntyy manuaalista työtä.

Toteutuksen mahdollistavia teknisiä ratkaisuja ei ole HUS Acamedicia lukuun ottamatta nykyisissä tietoturvalisissa käyttöympäristöissä. HUS Acamedicissakin auditointi rajoittuu parvioppimiseen. Algoritmin kautta haettua tietoa voi vuotaa esimerkiksi silloin, kun aineistossa on poikkeuksellisia havaintoja (esim. iältään yli 105-vuotias henkilö). Käyttötapausten toteuttamiseksi tarvitaan algoritmin auditointi varmistamaan yksityisyyden suoja ja varmistus, ettei algoritmi ei ole muutoin haitallinen.

5.2.4 Jatkokehitystarpeet

Tietoturvalisten käyttöympäristöjen tekninen toteutus tulee kehittää federoitua oppimista tukevaksi. Toteutus vaatisi luottamusverkoston laajentamista rajojen yli sekä erityisjärjestelyitä algoritmien tarkastamisen sekä tulosten anonymiteetin varmistuksen osalta. Tämän lisäksi tulee kehittää ennakkohyväksyntämenettely, jotta jokaisen yksittäisen tuloksen tarkastelua ei tarvitse suorittaa erikseen.

5.3 Hybridi federoidun oppimisen tutkimus (horisontaalinen ja vertikaalinen)

5.3.1 Kuvaus käyttötapauksesta

Tavoitteena on tehdä kansallinen tutkimus, jossa on mukana useita kansallisia rekisterinpitäjiä. Aineistojen muuttajat vaihtelevat datan lähteiden kesken.

5.3.2 Käyttötapausten merkitys

Esimerkkitutkimuksessa aineisto muodostetaan paljon palveluita käyttävistä yksilöistä. Aineisto sisältää useita kokonaisuuksia, joihin kuuluvat esimerkiksi hyvinvointialueet, Kela ja yksityinen kuntoutuslaitos.

5.3.3 Toteutusmahdollisuudet nykyisessä toimintaympäristössä

Teoriassa tällainen tutkimus on mahdollista toteuttaa. Toteutus voitaisiin tehdä joko useampana eri hankkeena tai yhtenä Findatan luvittamana hankkeena. Luvituksen osalta huomioitavaksi tulee aineistojen rekisterinpitäjyys tai mahdollinen yhteisrekisterinpitäjyys.

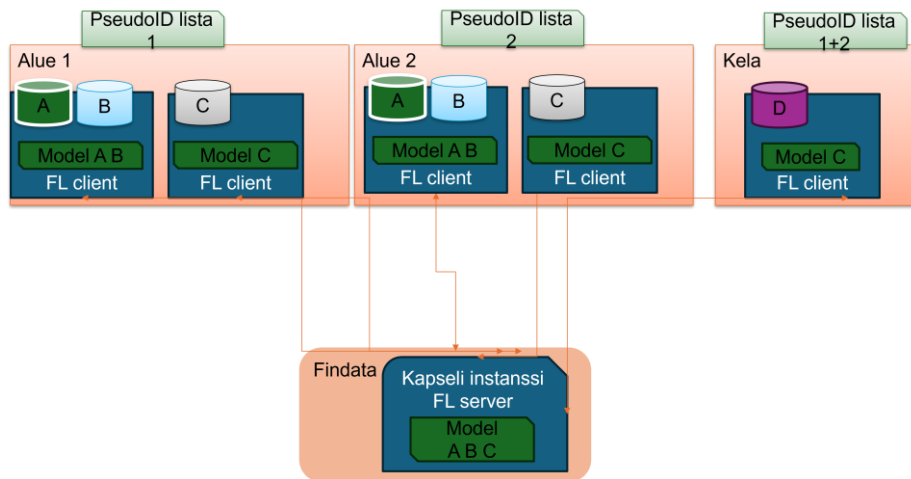
Toteutus edellyttää datan harmonisointia, esimerkiksi OMOP-mallinnusta sekä eri rekisterinpitäjien aineistojen muuttajien listaamista (vertikaalinen federoitu oppiminen). Lisäksi vertikaalinen data on harmonisoitava ja määriteltävä, miten tutkimushenkilöt linkitetään (esim. pseudonymisointi yhtenevillä koodilla).

Toteutus olisi mahdollista esimerkiksi Findatan standardoidun federoidun oppimisen verkon kautta (Kuva 2). Hybridi-mallissa jokaisella datan omistajalla olisi oma sertifioitu serveri, johon poimittu



tutkimusdata voidaan tuoda ja määritellä osaksi Findatan standardoituja federoidun oppimisen verkkoa. Tämän jälkeen Findatan Kapseli määritellään keskitetyksi laskentayksiköksi, jonka kautta federoidun oppimisen mallia ajetaan määrittelyssä verkostossa.

Tutkimuksen onnistuminen edellyttäisi standardoituja verkko- ja laskentaresursseja sekä yhtenäisiä on-demand deployment -menetelmiä. Lisäksi kaikkien hyvinvointialueiden ja muiden rekisterinpitäjien aineistojen tulee noudattaa yhtenäisiä OMOP-määrittelyjä, jotta analyysi voidaan toteuttaa luotettavasti.



Kuva 2. Havainnollistuskuva hybriditapauksen toteutuksesta

5.3.4 Jatkokehitystarpeet

Jotta tämä käytötapaus voitaisiin toteuttaa, tarvitaan aineistojen harmonisointia ja teknisten standardiratkaisujen kehittämistä (verkko ja deployment). Tutkimuksen jakaminen eri käyttöympäristöjen kesken voisi olla ratkaisu (federoitu laskenta). Esimerkiksi HUS Academicissa tehtävä tutkimus laajennettaisiin CSC:n ympäristössä tapahtuvaksi, jolloin voidaan hyödyntää CSC:n laskentakapasiteettia Academicin tietoturvalisessa käyttöympäristössä.

Määräys ei estä jakamasta laskentaa useampaan tietoturvaliseen käyttöympäristöön, kunhan aineisto on luvitettu asianmukaisesti, tietoturvalisesta käyttöympäristöstä ei siirry ulos henkilötietoja ja arviointilaitos on hyväksynyt ratkaisun toteutuksen. Toteutus voisi tapahtua luomalla virtuaalisia tietoturvalisista käyttöympäristökokonaisuuksia tietoturvalisten käyttöympäristöjen kesken. Uusien teknisten ratkaisujen myötä tuleva mahdollinen lisäauditoinnin tarve tulee huomioida.

6 Yhteenveto keskeisistä haasteista ja suositukset jatkokehitykseen

Käytötapausten tarkastelun kautta tunnistettiin tässä kappaleessa listatut haasteet 1–5.

1. Yhtenäisiä määrittelyjä federoidulle laskennalle ei ole ja siksi aiheeseen liittyy myös väärinymmärryksiä. Keskenään sekoittuvat niin kertaluontoinen federoitu analyysi kuin

iteratiivinen analyysi eli federoitu oppiminen. Lisäksi federoituun laskentaan viitataan joskus hajautettuna laskentana, mikä viittaa myös suurteholaskentaan eli täysin eri toimintaan.

Asiantuntijaryhmän suositukset jatkokehityksen osalta:

Seuraavat taustapaperin mukaiset määritelmät suositellaan otettavaksi käyttöön federoitua laskentaa koskevan keskustelun selkeyttämiseksi:

- **Federoitu laskenta** (eng. federated processing tai computing): yleistermi federoidulle laskennalle
- **Federoitu analyysi** (eng. federated analysis): kertaluontoinen federoitu laskenta
- **Federoitu oppiminen** (eng. federated learning): iteratiivinen federoitu laskenta.

2. Lainsäädännön suhde federoidun laskennan toteuttamiselle on aiheuttanut epäselvyyttä niin kansallisessa yhteistyössä kuin kansainvälisissä tietolupaan perustuvissa hankkeissa. Tässä taustapaperissa on pyritty selkeyttämään toisiolaista tulevia reunaehtoja. Toisilain mukaisen tietoluvan perusteella tapahtuvaa federoitua laskentaa on mahdollista toteuttaa niin, että henkilötietoa ei siirretä tietoturvalisesta käyttöympäristön ulkopuolelle. Tietojohdamisen osalta yksilötasoisia tietoja ei tule yhdistää muihin toisilain alaisiin aineistoihin. Tällöin siirrettävän tiedon tulee olla anonyymejä algoritmeja ja/tai tuloksia.

Toisilain vaatimukset eivät suoraan säätele federoitua laskentaa. Lähtökohtaisesti federoitu laskenta mahdollistaa sote-tietojen toisiokäytön niin, että yksilötasoisia aineistoja ei ole tarpeen yhdistää tai siirtää. Se on siten hyvänä pidettävä vaihtoehto koko aineiston koostamiselle yhteen paikkaan.

Federoidun laskennan toteuttamistavoissa on hyvä huomioida GDPR:stä tulevat rekisteröityjen oikeudet. Lisäksi on toivottavaa, että federoitu laskenta on mahdollista myös sovellettaessa EHDS-asetusta terveystietojen toisiokäytössä.

Asiantuntijaryhmän suositukset jatkokehityksen osalta:

- Federoitu laskenta on menetelmänä suositeltava ja sopisi joissakin tapauksissa ensisijaiseksi menetelmäksi. Sen mahdollistamiseksi tulisi selkeyttää myös tietoturvalisesta käyttöympäristöstä ulosviettävien välitulosten anonymisoinnin varmistamiseen liittyviä vaatimuksia (kts. myös suositus 4). Federoidun laskennan toteuttamisen tulisi jatkossakin olla mahdollista käyttötarkoituksesta riippumatta.
- Tietosuoja-asetuksen mukaisesti rekisteröityjen oikeuksien toteuttamisen prosessiin on tärkeä kiinnittää huomiota myös federoidun laskennan käyttötapauksissa. Kansalaisilla on oltava mahdollisuus saada tietoa tietojensa käytöstä myös silloin, kun heidän tietojensa hyödynnetään federoitua laskentaa käyttävässä hankkeessa. Tämän on oltava mahdollista myös silloin, kun toteuttajataholla ei ole kokonaisnäkymää henkilöistä, joiden tietoja tutkimuksessa hyödynnetään.
- 26.3.2025 voimaan astuneessa EHDS-asetuksessa sekä vuoteen 2027 mennessä annettavassa täytäntöönpanosäädöksessä tullaan säätelemään eurooppalaisten tietoturvalisesta käyttöympäristöjen vaatimuksista. Näissä vaatimuksissa on tärkeä huomioida federoidun laskennan mahdollistaminen.

3. Puutteellinen datan harmonisointi estää federoidun laskennan toteutuksen. Rajat ylittävässä datan hyödyntämisessä on huomioitava datan harmonisoinnin lisäksi myös sosioekonominen konteksti, jossa datat on tuotettu. Tämä ei kuitenkaan ole pelkästään federoidun laskennan toteutuksessa haasteena vaan kaikissa rajat ylittävissä projektissa.



Asiantuntijaryhmän suositukset jatkokehityksen osalta:

- Datan harmonisointityötä tulee jatkaa ja tukea niin kansallisesti kuin osana EHDS asetuksen toimeenpanoa.

4. Federoidun laskennan, erityisesti federoidun oppimisen, mahdollistamiseksi algoritmien välittämien tietojen anonymiteetti tulisi varmistaa silloin, kun tiedot lähtevät tietoturvalisen käyttöympäristön ulkopuolelle tai lupa ei kata tietojen yhdistämistä tietoturvalisten käyttöympäristöjen välillä. Federoidussa oppimisessa välitetyt tiedot eivät välttämättä ole anonymoijia, vaikka niissä ei siirretä alkuperäistä aineistoa.

Asiantuntijaryhmän suositukset jatkokehityksen osalta:

- Federoidussa oppimisessa hyödynnettävien algoritmien ja menetelmien hyväksyntäkriteerejä olisi selkeytettävä. Niiden osalta voitaisiin tehdä yhteistyötä tietoturvalisten käyttöympäristön tarjoajien sekä auditoijien välillä.
- Tietoturvalisesta käyttöympäristöstä ei saa tuoda ulos tietoja ilman kontrollointia. Tämän varmistamisesta vastaa käyttöympäristön palveluntarjoaja. Lisäksi laskentaympäristöjen välillä tapahtuvaa tiedonsiirtoa varten olisi hyvä ottaa käyttöön yhteiskäyttöinen tiedonsiirtoratkaisu, joka tehostaisi toimintaa ja parantaisi tietoturvalisuutta.

5. Tietoturvalisista käyttöympäristöistä HUS Acamedic on toteuttanut ensimmäisenä federoitua oppimista tukevan teknisen ratkaisun, jonka perusteella on mahdollista hyödyntää Swarm Learning -menetelmää. HUS Acamedicia lukuunottamatta teknisiä ja organisatorisia ratkaisuja ei kuitenkaan ole rakennettu kansallisiin auditoituihin tietoturvalisiin käyttöympäristöihin. Keskenään yhdistettyjen käyttöympäristöjen välillä ei myöskään ole käytäntöjä eikä standardoituja toimintatapoja federoitujen ympäristöjen verkoston hallinnoimiseksi.

Asiantuntijaryhmän suositukset jatkokehityksen osalta:

- Yhteistyötä tietoturvalisten käyttöympäristöjen tarjoajien välillä tulee jatkaa ja kehittää erilaisten teknisten ratkaisujen kartoittamiseksi ja vertailemiseksi sekä auditoijien hyväksymien teknisten ratkaisujen toteuttamiseksi.
- Osana tätä kehitystyötä tulee selvittää teknisiä vaihtoehtoja integraatoratkaisuille, jotka lainsäädännön vaatimukset huomioiden mahdollistavat yhteydet eri käyttöympäristöjen välillä sekä toisaalta käyttöympäristön ja käyttäjän hallinnassa olevan sovelluksen välillä.
- Osapuolten tulee voida varmistua käyttöympäristöt yhdistävien teknisten ratkaisujen tietoturvalista yhteistoiminnassa.