3.12.2020

Antti Honkela
Datatieteen apulaisprofessori
Tietojenkäsittelytieteen osasto
Helsingin yliopisto

Eduskunnan sosiaali- ja terveysvaliokunnalle

Asia: Asiantuntijapyyntönne HE 212/2020 vp

Lausunto Hallituksen esityksestä eduskunnalle laiksi sosiaali- ja terveydenhuollon asiakastietojen sähköisestä käsittelystä sekä eräiksi siihen liittyviksi laeiksi

Kiitän sosiaali- ja terveysvaliokuntaa mahdollisuudesta esittää lausunto sosiaali- ja terveydenhuollon asiakastietolaista. Oma asiantuntemukseni liittyy erityisesti tekniseen ja tilastolliseen tietosuojaan sekä tietojen hyödyntämiseen koneoppimista käyttävissä tekoälysovelluksissa, joten tarkastelen esitystä ensisijaisesti näistä näkökulmista.

Lakiesitys vaikuttaa kokonaisuutena hyvin valmistellulta, erityisesti perinteisiä asiakastietojen käsittelyyn tarkoitettuja tietojärjestelmiä koskevilta osin. Sen sijaan hyvinvointisovelluksia koskeva osa sisältää useita potentiaalisia ongelmia, joita käsittelen tarkemmin alla.

Esityksen perusteluissa (s. 145, 151) todetaan, että tietojärjestelmiin tallennettavia asiakastietoja ei ole mahdollista suojata pseudonymisoimalla, anonymisoimalla tai salaamalla, vaan suojauksen täytyy perustua hallinnollisiin ja pääsyä rajaaviin toimenpiteisiin. Olen samaa mieltä tästä periaatteesta. Joissain viimeaikaisissa keskusteluissa on Vastaamon tietovuodon jälkeen noussut esiin ajatus, että järjestelmiin pitäisi tallentaa ainoastaan anonymisoitua tai pseudonymisoitua tietoa. Tämä ei ole järkevää, koska anonymisointi tai pseudonymisointi vaatisi tiedon muokkaamista niin, että se ei enää olisi yhtä hyödyllistä potilaan hoidolle, vaan voisi vaarantaa potilasturvallisuuden. Lisäksi anonymisoinnin tai pseudonymisoinnin tekninen toteutus olisi hyvin haastavaa. Tähän ei ole tällä hetkellä olemassa käyttökelpoisia teknisiä työkaluja, ja ihmistyönä se kuormittaisi merkintöjen tekijöitä kohtuuttomasti.

Esitykseen sisältyvä omatietovaranto ja siihen liittyvät hyvinvointisovellukset ovat mielenkiintoisia uusia elementtejä. Yksinkertaiset kansalaisen hyvinvointitietoja omatietovarantoon tallentavat sovellukset ovat ehdottomasti kannatettavia. Yhtenäinen turvallinen ympäristö on merkittävä parannus nykytilaan, jossa esimerkiksi diabeetikon verensokeritiedot kulkevat terveydenhuoltoon ulkopuolisen kaupallisen palvelun kautta.

Esityksessä säädetään myös mahdollisuudesta luovuttaa asiakastietoja hyvinvointisovelluksille käyttäjän suostumuksella. Tätä koskeva sääntely on hyvin ylimalkaista. Nähdäkseni tähän voi sisältyä tietosuojariskejä, erityisesti jos usean käyttäjän luovuttamista tiedoista kerätään rekisteri. Näiden tietojen suojaamiseksi mahdollistettuja ja edellytettyjä toimia olisi syytä kehittää ja tarkentaa. Esitän alla joitain tarkempia ehdotuksia kehityskohteista.

Kokonaisuutena esitys toisi nähdäkseni lukuisia parannuksia nykytilaan, ja sitä kannattaa siksi edistää. Hyvinvointisovellusten sääntelyyn liittyy kuitenkin muutamia heikkouksia, jotka on syytä korjata.

Hyvinvointisovellusten sertifiointi

Hyvinvointisovellusten sääntelyssä ja sertifioinnissa pitäisi soveltaa riskiperusteista arviointia.

Ehdotuksessa luokitellaan 29 §:ssä kaikki hyvinvointisovellukset tiukimmin säädeltyyn luokkaan A, jolta edellytetään 30 §:n mukaan yleisen käyttötarkoituskuvauksen lisäksi yhteentoimivuustestausta ja 35 § mukaista sertifiointia. Nämä aiheuttavat sovellusten tekijöille kustannuksia ja siten toimivat esteenä uusien sovellusten syntymiselle. Este on erityisen merkittävä avoimen lähdekoodin sovellusprojekteille, joilla ei lähtökohtaisesti ole tulolähteitä kustannusten maksuun.

Ehdotuksessa tarkoitetut hyvinvointisovellukset voivat olla hyvin erilaisia. Sovelluksille, joille luovutetaan laajasti käyttäjien potilastietoja, olisi perusteltua soveltaa tiukempaa sääntelyä kuin sovelluksille, jotka ainoastaan kirjoittavat omatietovarantoon käyttäjän tietoja, joihin sovelluksella on joka tapauksessa pääsy.

Sovellusten käyttöoikeuksien jako olisi teknisesti helppoa toteuttaa ja valvoa: sertifioimattoman hyvinvointisovelluksen käyttöoikeudet omatietovarantoon voitaisiin helposti rajata vain sovelluksen itse kirjoittamiin tietoihin.

Tällainen muutos edistäisi lain tarkoituksen toteutumista, koska se todennäköisesti lisäisi omatietovarannon käyttöä helpottamalla huomattavasti siihen tietoa tuottavien hyvinvointisovellusten kehittämistä.

Mikäli tietojen luotettavuudesta terveydenhuollon käytössä ilmenee huolta, tämä voidaan helposti ratkaista vapaaehtoisella sertifioinnilla.

Ehdotus:
Lisätään 29 §:n hyvinvointisovellusten luokitteluun uusi luokka sovelluksille, joilla on pääsy ainoastaan sovelluksen itse omatietovarantoon kirjoittamiin tietoihin. Poistetaan näiden sertifiointipakko ja tehdään yhteensopivuustestaus mahdollisimman automaattiseksi ja kevyeksi.

Asiakastietojen luovutus hyvinvointisovelluksille

Hyvinvointisovellusten toteutuksen tietosuojaa voisi parantaa luomalla turvallisempia vaihtoehtoja asiakastietojen luovutukselle suoraan sovelluksille.

Ehdotuksessa säädetään mekanismeista, joilla asiakkaan suostumuksella voidaan luovuttaa asiakastietoja hyvinvointisovelluksille. Ajatus on tietojen laajemman hyödyntämisen kannalta hyvä, mutta toteutus jättää paljon kysymyksiä luovutettujen tietojen tietosuojasta.

Arkaluontoisten tietojen luovutuksen sijaan tai rinnalle olisi syytä harkita mekanismia, jossa esityksessä ehdotettujen rekisterien rinnalle luotaisiin tietoturvallinen käyttöympäristö tai tietoallas, jossa hyvinvointisovellukset voisivat käsitellä asiakastietoja turvallisesti siirtämättä kaikkia tietoja ulkopuolisiin järjestelmiin. Tietosuoja paranisi huomattavasti, kun tietoaltaasta siirrettäisiin ainoastaan käsittelyn tuloksena saatu rajoitetumpi tieto eikä muodostuisi ulkopuolisia asiakastietorekisterejä.

Ehdotus:
Valmistellaan hyvinvointisovellusten toteutukseen korkeamman tietosuojan tarjoavia mekanismeja, kuten ehdotettujen rekisterien rinnalla toimiva tietoturvallinen käyttöympäristö tai tietoallas, jossa hyvinvointisovellukset voisivat käsitellä niille luovutettuja asiakastietoja turvallisesti siirtämättä niitä heikommin valvottuihin ulkopuolisiin järjestelmiin.

Hyvinvointisovellusten kehittämiseen tarvittavan datan saatavuus

Esityksessä säädetään hyvinvointisovelluksista, jotka voivat käsitellä asiakkaiden asiakastietoja. Sovellusten määrittely on laaja, mutta olettaisin monien ajateltujen sovellusten perustuvan koneoppimiseen tai tekoälyyn, joka voisi ehdottaa asiakkaan tietojen perusteella tämän terveyttä ja hyvinvointia edistäviä toimenpiteitä.

Koneoppimiseen perustuvien sovellusten kehittäminen vaatii ns. opetusaineiston, joka koostuu suuren ihmisjoukon tiedoista. Ehdotuksen ja olemassa olevan sääntelyn perusteella on epäselvää, mistä sovellusten kehittäjät voisivat saada tällaisen opetusaineiston.

Koska sovelluksen kehittäminen ei ole asiakastietojen ensisijaista käyttöä, aineiston hankkiminen kuuluisi luontevimmin Lain sosiaali- ja terveystietojen toissijaisesta käytöstä (552/2019; ”toisiolaki”) alaan.

Toisiolaissa säädetään mahdollisuudesta käyttää asiakastietoja kehittämis- ja innovaatiotoimintaan. Tämä toteutuu Sosiaali- ja terveysalan tietolupaviranomaiselle (Findata) osoitetulla tietopyynnöllä, jonka vastauksena voidaan luovuttaa asiakastiedoista muodostettuja aggregoituja tilastotietoja.

Nämä aggregoidut tilastotiedot (tyypillisesti suuren ihmisjoukon tietojen keskiarvot tms.) eivät pääsääntöisesti ole käyttökelpoisia koneoppimisjärjestelmän opetusaineistona, jossa tarvittaisiin yksilötason aineistoa. *Nykyinen toisiolaki ei näin mahdollista hyvinvointisovellusten kehittämiseen tarvittavan opetusaineiston hankkimista*.

Toisiolain vaihtoehtona tarvittavan aineiston keruu voisi perustua käyttäjien suostumukseen. Tietojen luovutus hyvinvointisovellusten kehittäjille on mahdollista ehdotetun asiakastietolain 20 §:n ja 21 §:n nojalla. Tähän liittyvä sääntely on kuitenkin epämääräistä, mihin liittyy merkittäviä tietosuojariskejä.

Yhtenä konkreettisena uhkakuvana nousee esiin jotain näennäistä tai todellista hyötyä tarjoavat hyvinvointisovellukset, joiden todellisena tarkoituksena on kuitenkin kerätä käyttäjien asiakastietoja

sovelluksen kehittäjille. Vastaavalla periaatteella toimivia mobiilisovelluksia on markkinoilla runsaasti[1]. EU:n tietosuoja-asetus (GDPR) tarjoaa periaatteessa välineitä näihin puuttumiseen, mutta käytännössä se on toistaiseksi ollut tehoton.

Hyvinvointisovellusten käyttöä käyttäjien asiakastietojen laajamittaiseen keräämiseen rajoittaa lähinnä 35 §:n mukainen sovellusten sertifiointi. Tätä koskeva sääntely on ehdotetussa laissa ylimalkaista, koska siinä vaaditaan lähinnä, että sovellus täyttää ”olennaiset vaatimukset”, joita ei määritellä. Olennaisia vaatimuksia määriteltäessä on tärkeää pitää huolta, että niillä varmistetaan hyvinvointisovelluksille luovutettujen asiakastietojen riittävä tietosuoja. Muuten riskinä on heikommin säänneltyjen varjoasiakastietojärjestelmien syntyminen. Yhtenä mahdollisuutena voisi olla yllä esitetyn tietoaltaan kaltaiset tekniset ratkaisut, joilla ehkäistään tietojen vuotamisen riskiä. Yleisesti olisi syytä harkita säädöstä, joka vaatisi hyvinvointisovelluksille luovutettujen tietojen suojaamista pitkäaikaisessa säilytyksessä vastaavasti kuin muissa asiakastietojen käsittelyssä.

Hyvinvointisovellusten avulla tapahtuvan tiedonkeruun vaihtoehdoksi olisi syytä vahvasti harkita vaihtoehtoisia tapoja mahdollistaa tarvittavan opetusaineiston saaminen tietosuojaa heikentämättä, esimerkiksi toisiolakia kehittämällä.

Toisiolain haasteena on, että EU:n tietosuoja-asetus ei salli henkilötietoja sisältävien asiakastietojen käyttöä kehittämis- ja innovaatiotoiminnassa ilman suostumusta. Näin ainoaksi mahdollisuudeksi jää tietojen anonymisointi, jolloin ne eivät enää ole henkilötietoja. Toisiolain kirjaus aggregoidusta tilastotiedosta on tiukempi kuin mitä tietosuoja-asetus vaatisi. Tietosuoja-asetuksen näkökulmasta tietojen luotettava anonymisointi, jolla varmistetaan, ettei tietoja voi liittää tunnistettavaan tai tunnistettavissa olevaan yksilöön, riittäisi. Tietosuoja-asetus sallisi näin esimerkiksi anonymisoidun synteettisen aineiston käytön, mikä ei toisiolain perusteella ole selvää. Lisäksi tietosuoja-asetuksen puitteissa olisi mahdollista rakentaa koneoppimismallin opetukseen tarvittava pääsy tietoihin anonymisoivan ohjelmointirajapinnan kautta ns. differentiaalista tietosuojaa hyödyntäen.

Toisiolain lähtökohta on, että tietoja käsitellään ilman rekisteröidyn suostumusta. Tieteellisen tutkimuksen kohdalla tämä on välttämätöntä tutkimuksen vääristymisen välttämiseksi. Useimmat kansalaiset todennäköisesti suhtautuvat suopeasti tietojensa käyttöön tieteelliseen tutkimukseen. Tietojen käyttö kaupalliseen tarkoitukseen tähtäävän sovelluksen kehittämiseen mielletään kuitenkin usein eri tavalla. Ideaalinen ratkaisu ongelmaan olisi mekanismi, jolla kansalaiset voisivat joustavasti hallinnoida tietojensa käyttöä eri tarkoituksiin ja säätää vaaditun tietosuojan tasoa. Kansantalouden kannalta tuntuisi lisäksi järkevältä kehittää aineistojen hinnoittelua, joka nykyisin perustuu samaan hintaan niin kotimaisilta tutkijoilta kuin ylikansallisilta lääke- ja teknologiajäteiltä.

Ehdotus:
Täsmennetään 34 §:n hyvinvointisovellusten vaatimuksia niille luovutettujen asiakastietojen käsittelystä. Voidaan esimerkiksi harkita lisätä vaatimus, että hyvinvointisovelluksille luovutetuista asiakastiedoista muodostuvaa rekisteriä pitäisi käsitellä ja suojata yhtä huolellisesti kuin muissa asiakastietojärjestelmissä.

Kehitetään toisiolakia tai muita keinoja mahdollistamaan hyvinvointisovellusten kehittämiseen tarvittavan opetusaineiston saatavuus tietosuojaa kunnioittaen.

[1] https://www.forbrukerradet.no/out-of-control/