PSEUDONYMISOINTI, ANONYMISOINTI JA SUORIEN TUNNISTEIDEN KÄYTTÖ SOSIAALI- JA TERVEYSTIETOJEN TOISSIJAISESTA KÄYTÖSTÄ ANNETUN LAIN (522/2019) MUKAAN

Sosiaali ja terveystietojen toissijaisen käytön tietosuojan asiantuntijaryhmä (STM103:00/2019) linjaa seuraavaa pseudonymisoinnin ja anonymisoinnin toteuttamisesta sekä suoria tunnisteita sisältävän tiedon käytöstä sosiaali- ja terveystietojen toissijaisesta käytöstä annetun lain (552/2019, jäljempänä "toisiolaki") mukaisessa käsittelyssä. Toisiolaki pohjaa tietosuojan määrittelyjen osalta vahvasti EU:n yleiseen tietosuoja-asetukseen (2016/679).

Toisiolain mukaisia tietojenkäsittely-ympäristöjä on tunnistettu kaksi erilaista:

1) Tietoturvallinen käyttöympäristö, jonka osalta paljastumisriskiä tarkastellessa otetaan huomioon muutkin suojatoimet. Esimerkiksi tietojen tutkimuskäytön osalta tarkastellaan kokonaisuutta, jonka muodostavat tietoja käsittelevät henkilöt, tarkasteltava projekti ja aineisto, käytettävä käsittely-ympäristö sekä julkaistavien tulosten erillinen tarkastaminen. (Desai, Ritchie, Welpton 2016)

2) Muu käyttöympäristö, jossa voidaan käsitellä tietoturvallisessa käyttöympäristössä tuotettuja ja anonyymeiksi todettuja tuloksia tai tietopyynnön mukaisia aggregoituja tilastotietoja, joita voidaan myös jakaa julkisesti.

Tietoturvallisen käyttöympäristön määräykset kattavat seuraavat vaatimukset: tietoja käsittelevät henkilöt ovat vahvasti tunnistettuja, tietojen käyttötarkoitus on hyväksytty, käsittely-ympäristön tietoturvasta on huolehdittu ja käsiteltävistä aineistoista johdetut tulokset tarkastetaan paljastumisriskin näkökulmasta. Tällöin tietojen suojatoimia on käytössä useita ja käsiteltävät aineistot voivat olla laajoja ja sisältää yksityiskohtaista tietoa.

Tietopyynnön perusteella tuotettu aggregoitu tilastotieto tai tietoturvallisessa käyttöympäristössä käsitellyn aineiston perusteella johdetut tulokset voivat käyttötarkoituksesta riippuen tulla julkaistuiksi. Julkaistun aineiston osalta ainoa suojatoimi on aineiston anonymisointi.

Pseudonymisointi

EU:n yleisen tietosuoja-asetuksen 4 artiklan mukaan **pseudonymisoimisella** tarkoitetaan henkilötietojen käsittelemistä siten, että henkilötietoja ei voida enää yhdistää tiettyyn rekisteröityyn käyttämättä lisätietoja. Pseudonymisoitu aineisto on edelleen henkilötietoa, koska yksittäisen henkilön tunnistamisen mahdollisuutta ei ole kokonaan poistettu.

Toisiolain alaisten aineistojen osalta pseudonymisointi toteutetaan poistamalla aineistosta suorat tunnisteet (henkilötunnus, nimi, tarkka osoite) ja luomalla lähtökohtaisesti henkilötunnusta vastaava yksi pseudotunniste eri henkilöiden erottamiseksi ja aineistojen yhdistämistä varten. Pseudonymisoinnin toteutuksessa ei muuteta aineiston tilastollisia ominaisuuksia eli pseudonymisoinnissa ei tehdä karkeistamista tai käytetä muita tietojen tarkkuutta heikentäviä menetelmiä.

Anonymisointi

EU:n yleisessä tietosuoja-asetuksessa (2016/679) resitaalissa 26 viitataan " -- *anonyymeihin tietoihin eli tietoihin, jotka eivät liity tunnistettuun tai tunnistettavissa olevaan luonnolliseen henkilöön, tai henkilötietoihin, joiden tunnistettavuus on poistettu siten, ettei rekisteröidyn tunnistaminen ole tai ei ole enää mahdollista.*"

Ottaen huomioon kaikki keinot, joita voidaan kohtuudella olettaa käytettävän henkilön tunnistamiseen, anonyymin tiedon tulee täyttää seuraavat ominaisuudet[1,2]:

- Anonymisoidusta aineistosta ei voida tunnistaa suoraan tai välillisesti yksittäistä henkilöä.
- Yksittäisestä henkilöstä ei voida tehdä vain tätä henkilöä koskevia päätelmiä.
- Yksittäistä henkilöä koskevien tietojen yhdisteleminen anonymisoidun aineiston ja muun aineiston välillä ei ole mahdollista.
- Aineistoa ei ole mahdollista tai on kohtuuttoman vaikeaa palauttaa muotoon, josta yksittäinen henkilö on tunnistettavissa.

Aineistojen käsittely tietoturvallisessa käyttöympäristössä

Tietoturvallisessa käyttöympäristössä käsiteltävien aineistojen osalta toimitaan yleisten tietosuojaperiaatteiden ja EU:n yleisen tietosuoja-asetuksen 5 artiklan tietojen minimoinnin periaatteen mukaisesti. Toisiolain mukaisen tietoluvan perusteella käyttöön toimitetaan aina vain välttämättömät tiedot. Tämä tarkoittaa ettei aineistot sisällä lähtökohtaisesti suoria tunnisteita ja lisäksi aineiston sisältämä muuttujajoukko on tarkasti rajattu ja muuttujien tiedot aina niin yleisellä tasolla kuin kulloiseenkin käyttötarkoitukseen katsotaan riittäväksi. Tietoluvalla voidaan toimittaa myös hyvin karkealle tasolle käsiteltyä, liki anonyymiä tietoa, kun sen katsotaan olevan käyttötarkoitukseen riittävää.

Tietoturvallisen käyttöympäristön määritelmän mukaisten suojatoimien johdosta ympäristössä on mahdollista käsitellä joissakin tapauksissa myös **suoria tunnisteita sisältäviä** tietoja. Suorien tunnisteiden sisällyttäminen aineistoon vaatii tietolupaviranomaisen tai muun tietoluvasta päättävän tahon kokonaisharkintaa, jossa otetaan huomioon lainsäädäntö ja lupahakemuksessa osoitettu erityinen tarve tai käsiteltävän tiedon erityisluonne.

Aineistojen käsittely muualla kuin tietoturvallisessa käyttöympäristössä

Toisiolain 3 § mukaan tietopyyntö kohdistuu aggregoituun tilastotietoon. Saman pykälän mukaan aggregoidulla tilastotiedolla tarkoitetaan tilastomuotoista, luotettavasti anonymisoitua tietoa. Aggregoidussa tiedossa jokainen luku on muodostettu useammasta havainnosta. Tilastomuotoinen tieto tarkoittaa tässä yhteydessä ensiokäytön tiedoista koottua yksikkötason tietoa tai näistä aggregoitua tietoa.

Toisiolain 52 § mukaan tietoluvan nojalla luovutetuista aineistoista johdettujen tulosten julkaisemisessa tietolupaviranomainen varmistaa tietojen olevan anonyymejä. Tällaiset tiedot luovutetaan käyttötarkoitustaan vastaavasti, mutta ne voidaan myös julkaista. Julkaistuun tietoon kohdistuva ainoa suojatoimi on luotettava anonymisointi tai anonymiteetin tarkastaminen. Tämän aineistotyypin osalta paljastumisriskiä tulee arvioida ottaen huomioon kaikki keinot joita voidaan kohtuullisen todennäköisesti käyttää mainitun luonnollisen henkilön tunnistamiseen suoraan tai välillisesti[3] ja tiedon on oltava anonyymiä sekä aggregoidun tilastotiedon että yksikkötasoisen tiedon tapauksessa.

Kirjallisuutta

Desai T. Richie F. Welpton R. (2016) Five safes: designing data access for research. Economics Working Paper Series 1601, University of the West of England, Bristol https://uwe-repository.worktribe.com/output/914745 [05.05.2020]

[1] Tietosuojatyöryhmän (WP29) lausunto 05/2014 Anonymisoinnin tekniikoista, Kansainvälinen ISO - standardi ISO29100:2011
[2] EU:n yleinen tietosuoja-asetus (2016/679) 26 resitaali
[3] EU:n yleinen tietosuoja-asetus (2016/679) 26 resitaali