**5 Opinnäytetyö ja sensitiiviset aineistot**

Tässä osiossa perehdytään siihen, miten eettiset kysymykset otetaan huomioon sensitiivisten eli arkaluontoisten aineistojen anonymisoinnissa.

Tässä osiossa on kolme teemaa. Ensimmäisessä määritellään, mikä on henkilötieto. Toinen teema on [tunnisteellisuus](https://moodle.karelia.fi/mod/lesson/view.php?id=338174) ja kolmas tietojen [anonymisointi](https://moodle.karelia.fi/mod/lesson/view.php?id=338175" \o "Anonymisointi).

Kaikkiin kolmeen teemaan sisältyy monivalintatehtäviä. Tämän osion hyväksytty suoritus edellyttää kaikkien tehtävien läpäisemistä. Voit yrittää tehtäviä niin monta kertaa että saat ne oikein.

[Tämä osio pohjautuu Aineistonhallinnan käsikirjaan [verkkojulkaisu]. Tampere: Yhteiskuntatieteellinen tietoarkisto [ylläpitäjä ja tuottaja].](https://www.fsd.tuni.fi/aineistonhallinta/fi/) <https://www.fsd.tuni.fi/aineistonhallinta/fi/>

Mitä on henkilötieto?

**EU:n tietosuoja-asetuksen määritelmän mukaan henkilötiedoilla tarkoitetaan kaikkia tunnistettuun tai tunnistettavissa olevaan luonnolliseen henkilöön liittyviä tietoja.** Tunnistettavissa olevana pidetään luonnollista henkilöä, joka voidaan suoraan tai epäsuorasti tunnistaa erityisesti tunnistetietojen, kuten nimen, henkilötunnuksen, sijaintitiedon, verkkotunnistetietojen taikka yhden tai useamman hänelle tunnusomaisen fyysisen, fysiologisen, geneettisen, psyykkisen, taloudellisen, kulttuurillisen tai sosiaalisen tekijän perusteella. (EU:n tietosuoja-asetus, [artikla 4](http://www.privacy-regulation.eu/fi/4.htm), kohta 1.) Määritelmän mukaisesti henkilötiedot tutkimusaineistoissa eivät rajaudu vain tutkittavia koskeviin tietoihin. Tutkimusaineistoihin voi sisältyä tunnistetietoja tutkittavien lähipiiristä tai muista kolmansista henkilöistä. Myös heitä tunnistettavasti käsittelevät tiedot ovat aina henkilötietoja.

Henkilötiedon sisältöä tai luonnetta ei ole rajattu. Kaikki luonnolliseen henkilöön liittyvät tiedot voivat olla henkilötietoja. Tiedot voivat olla väitteitä, mielipiteitä tai arvoarvostelmia. Ne voivat olla objektiivisia tai subjektiivisia. Niiltä ei edellytetä totuutta tai todennettavuutta. Tiedot voivat koskea yksityiselämää, perhe-elämää, terveydentilaa, fyysisiä ominaisuuksia, ammatillista toimintaa tai taloudellista ja sosiaalista käyttäytymistä.

Henkilötietoja ovat sellaiset tiedot, joiden perusteella henkilö voidaan tunnistaa suoraan tai välillisesti esimerkiksi yhdistämällä yksittäinen tieto johonkin toiseen tietoon, joka mahdollistaa tunnistamisen. Henkilö voidaan tunnistaa esimerkiksi nimen, henkilötunnuksen tai jonkin hänelle tunnusomaisen tekijän perusteella.

Esimerkkejä henkilötiedoista:

* nimi
* kotiosoite
* sähköpostiosoite, kuten etunimi.sukunimi@yritys.com
* puhelinnumero
* henkilökortin numero
* auton rekisterinumero
* paikannustiedot (esim. matkapuhelimen paikannustiedot)
* IP-osoite
* potilastiedot
* lemmikin eläinlääkäritiedot
* isoisoisovanhempien perinnöllisiä sairauksia koskevat tiedot.

Tunnisteellisuus

Milloin tieto on tunnisteellista?

Tieto on tunnisteellista, jos sen perusteella voidaan tunnistaa yksittäinen henkilö tai havaintorypäs, kuten samaan kotitalouteen kuuluvat henkilöt. Tunnistaminen voidaan tehdä yhden tai useamman henkilölle tunnusomaisen fyysisen, psyykkisen, taloudellisen, kulttuurisen tai sosiaalisen tekijän perusteella. Ei-ihmisiä käsittelevä tieto voi olla myös tunnisteellista, jos siinä on henkilöihin liittyviä tietoja (Elliot, Mackey, O´Hara & Tudor 2016). Esimerkiksi palolaitosten tiedot hälytyksistä voivat sisältää osoitetietoja.

Tietoja, jotka yksin riittävät tunnistamaan henkilön ovat henkilön koko nimi, henkilötunnus, henkilönimen mukainen sähköpostiosoite ja biometriset tunnisteet (sormenjälki, kasvokuva, ääni, silmän iiris, kämmenen muoto, käsin tehty allekirjoitus). Näistä käytetään nimitystä *suorat tunnisteet* tai välittömät tunnisteet.

Yksittäiset tiedot, joiden avulla henkilön voi kohtuullisen helposti tunnistaa ovat esimerkiksi postiosoite, puhelinnumero, auton rekisteri, henkilön julkaiseman teoksen viitetiedot, muu kuin henkilönimenmukainen sähköpostiosoite ja henkilöstä tunnistetietoja sisältävän verkkosivuston osoite, harvinainen ammattinimike, hyvin harvinainen sairaus tai vain yhdelle kerrallaan annettu asema (esimerkiksi puheenjohtajuus yhdistyksessä). Myös harvinainen tapahtuma voi olla avain henkilön tunnistamiseen. Tietoarkistossa näistä käytetään nimitystä *vahvat epäsuorat tunnisteet*.

Vahvoiksi epäsuoriksi tunnisteiksi luetaan Tietoarkistossa myös yksilöivät koodit, joiden avulla rajatulla henkilöiden joukolla on mahdollisuus tunnistaa henkilö yksiselitteisesti. Tällaisia ovat esimerkiksi opiskelijatunnus, vakuutusnumero, tilinumero, tietokoneen IP-osoite ja vastaavat.

*Epäsuoriksi tunnisteiksi* luetaan tiedot, jotka yksin eivät riitä tunnistamiseen, mutta yhdistettynä voivat mahdollistaa henkilön tunnistamisen. Epäsuorista tunnisteista käytetään englanniksi termejä *quasi-identifier* ja *indirect identifier*. Taustamuuttujia ja epäsuoria tunnisteita ovat esimerkiksi sukupuoli, ikä, koulutus, ammattiasema, pääasiallinen toiminta/työmarkkina-asema, sosioekonominen asema, kotitalouden koostumus, tulot, siviilisääty, kieli, kansallisuus, etninen tausta, työpaikka tai koulu ja asuinaluetta koskevat muuttujat. Asuinaluetta koskevat epäsuorat muuttujat kuuluvat niin sanottuihin aluemuuttujiin, joita voivat olla esimerkiksi postinumero, kaupunginosa, kunta, maakunta, seutukunta ja suuralue.

Myös päivämäärä voi olla epäsuora tunniste. Tavallisin päivämäärä epäsuorana tunnisteena on syntymäaika, mutta myös kuolinpäivämäärät tai uutiskynnyksen ylittäneet tapahtumapäivämäärät ja vastaavat voivat olla epäsuoria tunnisteita tutkimusaineistoissa yhdistettynä muihin tietoihin. Lääke- ja terveystieteen aineistoissa epäsuoria tunnisteita muihin tietoihin yhdistettynä voivat toisinaan olla myös näytteenottoja ja muita hoitotapahtumia koskevat päivämäärät.

Myös *pseudonyymit aineistot* katsotaan henkilötiedoiksi. Sellaisia ovat esimerkiksi pitkittäistutkimusten aineistot, joissa analysoitavissa aineistoissa on havaintotunnus henkilötunnuksen sijaan ja tutkimusryhmällä on hallussa koodi, jonka avulla tunnisteeton aineisto voidaan yhdistää tutkittaviin henkilöihin.

Tunnisteellisten aineistojen käsittely

Tunnisteellisia aineistoja voi käyttää tieteelliseen tutkimukseen silloin, kun se on tarkoituksenmukaista, suunniteltua, asiallisesti perusteltua ja tietojen käsittelyyn on laillinen käsittelyperuste (esimerkiksi tutkittavan suostumus tai yleisen edun mukainen tutkimus).

Tutkittavien näkökulmasta tunnisteellisen aineiston käsittely muodostaa riskin, jos heitä koskevat luottamukselliset tiedot vuotavat ulkopuolisille, esimerkiksi tutkittavan lähipiirille, työnantajalle tai viranomaisille. Tämän vuoksi tunnisteellisten tutkimusaineistojen käsittelyn tulee olla suunnitelmallista ja huolellista. Tutkittavien yksityisyyden suojaa ei saa vaarantaa esimerkiksi aineiston huolimattomalla säilyttämisellä tai suojaamattomilla sähköisillä siirroilla. Henkilötietojen käsittelyyn voi soveltaa tässä käsikirjassa selitettyjä suojatoimia kuten minimointia, pseudonymisointia ja anonymisointia. [Anonymisointi](https://moodle.karelia.fi/mod/lesson/view.php?id=338175" \o "Anonymisointi) on yksi mahdollisuus avata aineistot uudelleenkäyttöön. Tarvittaessa aineistojen suojaamiseen voi käyttää myös käyttöoikeuksien hallinnollisia ja teknisiä tietoturvaratkaisuja.

» [Lisätietoa tietoturvasta](https://www.fsd.uta.fi/aineistonhallinta/fi/fyysinen-sailytys.html#tietoturva)

Termit haltuun

**Anonyymi tieto:** Yksittäinen havainto (henkilö) ei ole kohtuullisin keinoin tunnistettavissa annettujen tietojen perusteella tai tietoja muihin tietoihin yhdistämällä. Täysin anonyymiä tietoa ei ole olemassa, mutta hyvin toteutetuin menetelmin voi päästä tulokseen, jossa kohtuullisin keinoin ei voi yksittäisiä henkilöitä tunnistaa. *Anonymisoinnilla* kuvataan niitä moninaisia tapoja ja välineitä, joilla anonyymi tieto saavutetaan.

**Pseudonyymi tieto:** Yksittäinen havainto ei ole tunnistettavissa pseudonymisoitujen tietojen perusteella ilman ulkopuolisia lisätietoja. *Pseudonymisointi* on aineiston tunnisteellisten tietojen poistamista tai korvaamista peitetiedolla tai koodeilla, jotka prosessin jälkeen säilytetään erillään aineistosta ja suojataan organisatorisesti ja teknisesti. Aineisto on pseudonyymi niin pitkään kuin erillään oleva tieto tuhotaan.

**De-identifiointi (de-identification):** Tunnisteellisten tietojen poistaminen tai muokkaaminen yksittäisen havainnon tunnistamisen ehkäisemiseksi datatiedoston sisällä. Usein de-identifioinnilla tarkoitetaan suorien tunnisteiden poistamista tai peittämistä (Elliot ym. 2016).

**De-[anonymisointi](https://moodle.karelia.fi/mod/lesson/view.php?id=338175" \o "Anonymisointi):** Anonyymiksi luokitellun tiedon saattaminen tunnisteelliseksi yhdistämällä tietoja muista lähteistä saataviin tietoihin. Jos anonyymi aineisto onnistutaan de-anonymisoimaan, aineisto joko ei ole ollut alun perin anonyymi tai käytettävissä oleva teknologia on kehittynyt tai henkilöistä on saatavissa muualta aiempaa enemmän tietoja. Anonyymisoinnin kestävyyttä tuleekin arvioida määräajoin (niin sanottu *jäännösriskin arviointi*).

**Erityiset henkilötietoryhmät:** Tietosuojalainsäädännön mukaisia henkilötietoja, joista ilmenee rotu tai etninen alkuperä, poliittisia mielipiteitä, uskonnollinen tai filosofinen vakaumus tai ammattiliiton jäsenyys, geneettiset tiedot, henkilön tunnistamista varten käsitellyt biometriset tiedot, terveyttä koskevat tiedot ja seksuaalista käyttäytymistä ja suuntautumista koskevat tiedot.

**Minimointi:** Henkilötietoja kerätään alun alkaen vain siinä määrin kuin on välttämätöntä tutkimuksen toteuttamiseksi. Henkilötietoja ei saa kerätä vain siltä varalta, että ne saattavat olla hyödyllisiä. Henkilötietojen keräämiselle tulee olla suunniteltu tutkimuksellinen tarve.

**Säilytyksen rajoittaminen:** Tutkimuksen toteuttamiselle tarpeettomat henkilötiedot poistetaan heti, kun se on mahdollista. Esimerkiksi aineiston keruuvaiheessa tarvitut nimitiedot, osoitteet ja vastaavat tunnisteet hävitetään heti, kun ne eivät ole enää välttämättömiä tutkimuksessa. Samoin tietojen yhdistämiseen tarvittu henkilötunnus voidaan hävittää, kun sitä ei enää tarvita. Säilytyksen rajoittaminen vähentää tunnisteellisten aineistojen käsittelyyn sisältyviä riskejä.

Milloin tieto on anonyymiä, entä pseudonyymiä?

Tieto on anonyymiä eli tunnisteetonta, jos tunnusomaiset piirteet (esimerkiksi epäsuorat tunnisteet yhdistettynä) koskevat samanlaisina useampaa henkilöä ja jos katsotaan, että henkilöä ei voida tunnistaa huomioiden kohtuullisesti toteutettavissa olevat toimenpiteet.

EU:n tietosuoja-asetuksessa anonyymi tieto määritellään toiminnallisesti:

Jotta voidaan määrittää, onko luonnollinen henkilö tunnistettavissa, olisi otettava huomioon kaikki keinot, joita joko rekisterinpitäjä tai muu henkilö voi kohtuullisen todennäköisesti käyttää mainitun luonnollisen henkilön tunnistamiseen suoraan tai välillisesti, kuten kyseisen henkilön erottaminen muista. Jotta voidaan varmistaa, voidaanko keinoja kohtuullisen todennäköisesti käyttää luonnollisen henkilön tunnistamiseen, olisi otettava huomioon kaikki objektiiviset tekijät, kuten tunnistamisesta aiheutuvat kulut ja tunnistamiseen tarvittava aika sekä käsittelyajankohtana käytettävissä oleva teknologia ja tekninen kehitys.

Lähde: Eu: tietosuoja-asetus, [johdantolause 26](http://www.privacy-regulation.eu/fi/r26.htm)

Anonyymistä aineistosta ei voi tunnistaa yksittäisiä tutkittavia tai kolmansia henkilöitä esimerkiksi epäsuorien tunnisteiden avulla tai yhdistämällä aineistoon muualta saatavia tietoja. Anonyymiin aineistoon ei voi myöskään yhdistää samoja tutkittavia koskevia uusia tietoja ja sen tulee olla peruuttamaton, jotta voidaan puhua anonyymistä aineistosta.

Pseudonyymiksi tiedoksi kutsutaan tietoa, jonka perusteella henkilöä ei voi tunnistaa ilman erillään säilytettäviä lisätietoja. Kun aineisto pseudonymisoidaan, aineiston tunnisteelliset tietosisällöt korvataan joko alkuperäisistä arvoista johdetuilla tai niistä riippumattomilla tiedoilla niin, ettei henkilö ole enää tunnistettavissa. Lisäksi tieto alkuperäisistä arvoista ja muutettujen arvojen muodostamistavoista pidetään erillään pseudonyymistä datasta organisatorisesti ja teknisesti. Organisatorisilla toimenpiteillä tarkoitetaan tietojen suojattua fyysistä käyttöympäristöä ja hallinnollisesti rajattua ja valvottua käyttöoikeutta. Teknisillä toimenpiteillä viitataan tietoturvallisiin tallennusratkaisuihin, kuten tietojen salausjärjestelmiin. (Tarhonen 2016.)

Tieto ei ole pseudonyymiä, jos yksittäinen henkilö on tunnistettavissa aineiston tietojen perusteella ilman erillään pidettäviä lisätietoja. Näin voi olla esimerkiksi tilanteessa, jossa aineiston epäsuorat muuttujat ja harvinaiset havainnot mahdollistavat tunnistamisen, vaikka henkilötunnus ja muut suorat tunnisteet säilytetään suojatusti erillään. Pseudonymisointi ei ole onnistunut myöskään, jos ulkopuolinen pystyy valituista pseudonyymeistä päättelemään alkuperäiset arvot. Näin voi käydä, jos osa alkuperäisistä tunnistetiedoista vain peitetään, esimerkiksi "Arja Kuula-Luumi" muutetaan muotoon "Arxx Kuxx-Luxx" tai henkilötunnus 123456-789E muutetaan muotoon 123456-XXXX. Henkilötunnuksen alkuosa kertoo täsmällisen syntymäpäivän ja vuoden ja on siten jo itsessään varsin vahva epäsuora tunniste. Pseudonymisoinnissa tulee tavoitella ratkaisuja, joiden perusteella ulkopuolinen ei mitenkään pysty päättelemään henkilötietoja.

Pseudonyymistä aineistosta tulee anonyymi, kun erillään säilytettävät tunnistetiedot (koodiavain, henkilötiedot ja tiedot muutettujen arvojen muodostamistavoista) hävitetään. Jos pseudonymisoidun aineiston erillään säilytettäviä henkilötietoja ei voi hävittää, aineistosta voi tehdä anonyymin tuhoamalla koodiavain ja muutettujen arvojen muodostamistiedot sekä järjestämällä aineisto uudella tavalla, esimerkiksi arvottujen uusien havaintotunnusten mukaan. Aineisto on anonyymi, jos sitä ei voi kohtuullisin keinoin enää yhdistää alkuperäisiin henkilötietoihin.

Esimerkiksi pitkittäistutkimuksen aineisto on tunnisteellinen niin kauan kun tutkimusryhmällä on hallussa koodiavain tutkittavien henkilötietoihin. Aineistosta ei tule myöskään anonyymiä, vaikka alkuperäinen koodiavain olisi koodattu kahteen kertaan (ns. kaksoiskoodaus). Koodaus ja kaksoiskoodaus ja ylipäänsä aineistojen pseudonymisointi ovat kuitenkin hyödyllisiä suojausmenetelmiä, kun halutaan estää tunnisteellisten tietojen käyttö analyyseissä. Koodaus ja kaksoiskoodaus kuuluvat esimerkiksi lääketieteen tutkimuskäytäntöihin.

Lisätietoa pseudonymisoinnista ja sen toteuttamiseen käytettävistä menetelmistä, voi lukea esimerkiksi Euroopan unionin verkko- ja tietoturvaviraston (ENISA) raportista: [Recommendations on shaping technology according to GDPR provisions. An overview on data pseudonymisation](https://www.enisa.europa.eu/publications/recommendations-on-shaping-technology-according-to-gdpr-provisions" \o "Recommendations on shaping technology according to GDPR provisions - An overview on data pseudonymisation) (November 2018).

https://www.enisa.europa.eu/publications/recommendations-on-shaping-technology-according-to-gdpr-provisions

Minimointi eli miten kerätä aineisto niin, ettei se sisällä turhia tunnisteita?

Ydinajatus anonymisointityön helpottamisessa on alkujaan välttää turhien, yksityiskohtaisten tai tutkimuksen kannalta merkityksettömien tietojen keräämistä. Huolellisessa suunnittelussa punnitaan, mitä taustatietoja henkilöstä tarvitaan tutkimukseen ja millä tarkkuudella. Lisäksi kysymysten muotoilulla vaikutetaan vastausten yksityiskohtaisuuteen. Määrällisissä tutkimuksissa kerättäviä tietoja pystyy kontrolloimaan helpommin verrattuna laadulliseen tutkimukseen valmiiksi luokiteltujen vastausvaihtoehtojen vuoksi.

Tulee muistaa, että myös tietosuoja-asetus kieltää tarpeettomien henkilötietojen keruun. Ole erityisen tarkka tietojen keruussa, jos keräät erityisiin henkilötietoryhmiin kuuluvia henkilötietoja, joista ilmenee esim. etninen alkuperä, poliittinen mielipide, uskonnollinen tai filosofinen vakaumus, terveystietoja tai seksuaalista käyttäytymistä ja suuntautumista koskevia tietoja.

Alla on vinkkejä aineiston keruuseen määrällisille ja laadullisille aineistoille minimoinnin näkökulmasta.

**Kvantitatiiviset aineistot**

* Älä kerää tietoja, jotka ovat harvinaisia kohdejoukossa. Jos harvinaisesta asiasta halutaan tietoa, tieto kannattaa kysyä valmiiksi luokiteltuna tai karkeistettuna, niin että harvinaisten joukossa on useampi harvinainen asia. Käytä luokittelussa hyödyksi tutkimuksessa yleensä käytettyjä henkilöluokituksia, joista voi katsoa mallia esim. [Tilastokeskuksen sivuilta](https://www.stat.fi/meta/luokitukset/index_henkilo.html) .
* Jos haluat minimoida henkilötiedot, älä käytä avokysymyksiä. Et voi kontrolloida niiden sisältöä. Jos silti haluat käyttää avokysymyksiä, mieti miten kysymyksen asettelulla voi vaikuttaa saataviin tietoihin. Älä kysy myöskään taustatietoja, kuten ammattia tai koulutusta avokysymyksillä, vaan luokiteltujen muuttujien avulla. Pohdi tarkkaan avokysymyksellä saatavan tiedon hyödyllisyyttä ja käytettävyyttä tulevaisuuden tutkimuksessa. Esim. kysymyspatteristojen jälkeen tulevat muu, mikä -tyyppiset avovastaukset tuottavat usein ainutlaatuisia tai harvinaisia tietoja, joiden perusteella henkilön voi tunnistaa.
* Luokiteltuna kannattaa kerätä muun muassa seuraavia tietoja: ammatti, tulot, ammattiasema, koulutus, kansalaisuus ja lasten lukumäärä. Arvoja ei tarvitse aina luokitella kokonaan vaan voi käyttää myös ääriarvojen luokittelua (esimerkiksi lasten lukumäärä: 0; 1; 2; 3; 4-tai enemmän).

**Kvalitatiiviset aineistot**

* Jos aineistonkeruutilanne ja sen luonne antaa mahdollisuuden, voit ystävällisesti muistuttaa tutkittavia ennen materiaalin keruuta, että henkilöiden nimiä, tarkkoja päivämääriä, työpaikkojen nimiä ja etenkin kolmansiin (ulkopuolisiin) henkilöihin liittyviä yksityiskohtaisia tietoja kannattaa välttää. Ihmiset ovat yllättävän hyviä anonymisoimaan kokemuksiaan ja tapahtumia.
* Mieti tarkkaan mitä taustatietoja keräät tutkittavasta ja miten sen toteutat. Taustatiedot voi kerätä strukturoidun lomakkeen avulla, jotta vältetään haastatteluiden alussa olevat haastateltavan vapaamuotoiset (ja usein tunnistettavat) esittelyt itsestään. Käytä luokittelussa hyödyksi tutkimuksessa yleensä käytettyjä henkilöluokituksia, joista voi katsoa mallia esim. [Tilastokeskuksen sivuilta](https://www.stat.fi/meta/luokitukset/index_henkilo.html) .
* Älä kysy haastatteluissa tai kirjoituskutsuissa tarkentavia kysymyksiä asioihin, jotka tuottavat hyvin todennäköisesti tietoa, jotka vaativat paljon anonymisointia (Ei näin: "Kertoisitko missä työpaikoissa äitisi ja isäsi ovat olleet töissä elämänsä aikana...")

Anonymisointi

Anonymisoinnin lähtökohdat

Anonymisointi on kansainvälisessä alan kirjallisuudessa laaja yläkäsite, joka pitää sisällään erilaisia lähestymistapoja, kuten saatavuuden hallintaa tai tilastollista lähestymistapaa (Elliot ym. 2016). Tietosuoja-asetuksen määritelmän mukaisesti saatavuuden hallinta on suojatoimi, mutta se ei ole anonymisointia. Tässä keskitymmekin dataan kohdistuviin anonymisointikeinoihin, joilla pyritään poistamaan kaikki tiedot, jotka mahdollistavat henkilön tunnistamisen.

Tutkimusaineiston anonymisointiin ei ole olemassa yhtä valmista kaikkiin aineistoihin soveltuvaa menettelytapaa. Anonymisointi tulee suunnitella aina aineistokohtaisesti ottaen huomioon aineiston ominaisuudet, käyttöympäristö ja käytettävyys.

Aineiston ominaisuuksilla viitataan esimerkiksi aineiston ikään, arkaluonteisuuteen, vastaajajoukon kokoon ja sisällön yksityiskohtaisuuteen (Elliot ym. 2016). Käyttöympäristöllä tarkoitetaan aineiston käyttökontekstia, joka on aikaan ja paikkaan sidottua: ketkä dataa käyttävät ja missä? Mitä ulkopuolisia tietoja on saatavilla sillä hetkellä? Käyttöympäristöllä tarkoitetaan myös [fyysistä säilyttämistä](https://www.fsd.uta.fi/aineistonhallinta/fi/fyysinen-sailytys.html#tietoturva). Käytettävyyden arvioinnissa pohditaan, miten anonymisoinnin saa tehtyä niin, että aineisto olisi edelleen käyttökelpoinen tilastollisessa tutkimuksessa tai laadullisessa tutkimuksessa.

Anonymisointi kannattaa suunnitella huolellisesti ja dokumentoida tehdyt toimenpiteet perusteluineen. Varaa anonymisointiin myös aikaa. Anonymisointi kannattaa huomioida jo aineistonkeruuvaiheessa, sillä huolellisella suunnittelulla anonymisointiin kuluvia resursseja voi säästää huomattavasti. Tutkijan kannattaa miettiä anonymisoinnista etukäteen seuraavat asiat:

* Varmista, että aineistonkeruuvaiheessa tietoja kerätään minimoinnin periaatteen mukaisesti.
* Päätä, kuka anonymisoinnin suunnittelee ja toteuttaa sekä missä vaiheessa ne tapahtuvat.

Anonymisointisuunnitelma

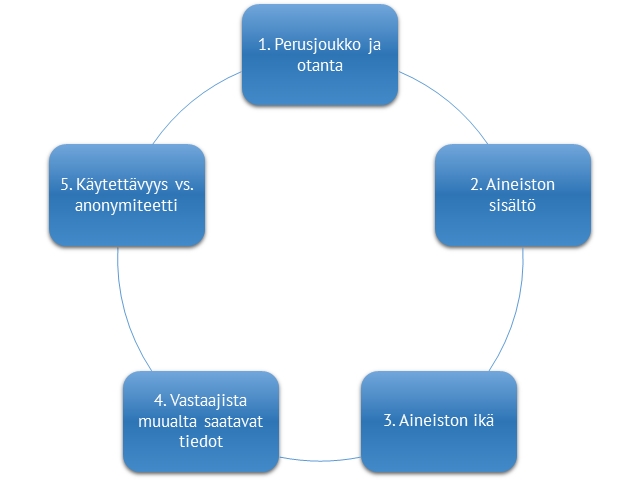
Anonymisointisuunnitelmaan kuvataan anonymisointitoimet ja arvioidaan henkilöiden paljastumisriskiä. Samalla suunnitelma toimii dokumenttina siitä, miten aineistoa on muokattu. Tiedot ovat tärkeitä esimerkiksi aineiston jatkokäyttöön arkistoimista varten tai kollegoille tutkimusyhteistyötä tehdessä. Myös tietosuoja-asetus edellyttää dokumentointia henkilötietojen käsittelyä koskevista päätöksistä. Suunnitelman teko voi alkaa jo aineistonkeruuvaiheessa.

Anonymisointisuunnitelmaan on hyvä koota seuraavia tietoja: suunnitelman tekijät, anonymisoinnin toteuttajat, anonymisointiin vaikuttavien aineiston ominaisuuksien kartoitus, vastaajien paljastumisriskin arvionti ja anonymisointitoimenpiteet perusteluineen. Lopussa voi myös arvioida henkilöiden tunnistamisen mahdollisuutta anonymisoinnin jälkeen ja tarvetta jäännösriskin arviointiin tulevaisuudessa.

Alla on esitetty Tietoarkiston esimerkit kvantitatiivisille ja kvalitatiivisille aineistoille tehdyistä anonymisointisuunnitelmista. Lisäksi saatavilla on anonymisointisuunnitelman tekoa helpottava mallipohja.

1. [Esimerkki kvantitatiivisen aineiston anonymisointisuunnitelmasta](https://www.fsd.uta.fi/aineistonhallinta/fi/dokumentit/anf-kvanti.pdf) (PDF) <https://www.fsd.tuni.fi/fi/palvelut/aineistonhallinta/anf-kvanti.pdf>
2. [Esimerkki kvalitatiivisen aineiston anonymisointisuunnitelmasta](https://www.fsd.uta.fi/aineistonhallinta/fi/dokumentit/anf-kvali.pdf) (PDF) https://www.fsd.tuni.fi/fi/palvelut/aineistonhallinta/anf-kvali.pdf
3. [Anonymisointisuunnitelman mallipohja](https://www.fsd.uta.fi/aineistonhallinta/fi/dokumentit/anf-pohja.pdf) (PDF) <https://www.fsd.tuni.fi/fi/palvelut/aineistonhallinta/anf-pohja.pdf>

Anonymisointi aloitetaan kartoittamalla aineiston ominaisuuksia, jotka esitellään seuraavassa osiossa. Tärkeimmät suunnittelussa huomioon otettavat seikat määrällisissä ja laadullisissa tutkimusaineistoissa on esitetty seuraavassa kuviossa.



**1. Perusjoukko ja otanta**

Ketkä olivat tutkimuksen kohteena ja miten otanta tehtiin? Kuinka moni lopulta päätyi aineistoon perusjoukosta? Mitä perusjoukon ominaisuuksista tiedetään jo etukäteen, esim. sukupuoli- ja ikäjakauma? Yhdistääkö kohdejoukkoa jokin harvinainen ilmiö?

Perusjoukko on aineistonkeruun kohteena oleva populaatio/asia ja otantamenetelmä kuvaa havaintoyksiköiden valikoitumista aineistoon. Suunniteltaessa anonymisointia arvioidaan aluksi aineiston perusjoukkoa ja otantamenetelmää, jotka voivat antaa itsessään harvinaisia tai ainutlaatuisia tietoja tutkimukseen osallistuneista henkilöistä. Perusjoukko voi olla tarkoin määritelty ja ulkoapäin tunnistettava ryhmä, kuten Tampereen kunnanvaltuutetut vuonna 2009 tai satunnainen ei-ulkoapäin määriteltävissä oleva ryhmä, kuten seksuaalista häirintää kokeneet suomalaiset. Perusjoukon koko on tärkeä, sillä mitä pienempi perusjoukko tai tutkittava ilmiö ovat, sitä suurempi mahdollisuus henkilö on tunnistaa.

Perusjoukon ja otannan osalta tulee pohtia, kuinka satunnaisia kohdejoukkoon kuuluvat ja lopulta tutkimukseen valikoituneet ovat suhteessa laajempaan mittakaavaan, kuten alueen väestöön. Seuraavassa on esitetty joitain esimerkkejä.

Kokonaisaineisto: Tutkimuskutsu lähetetään jokaiselle perusjoukkoon kuuluvalle, kuten kaikille suomalaisille alle 1-vuotiaiden keskoslasten vanhemmille tai tietyn kunnan täysi-ikäisille kuntalaisille. Näin voidaan jo etukäteen tietää henkilön sisältyvän mahdollisesti aineistoon.

Satunnaisotanta: Perusjoukkoon kuuluvalla on lähtökohtaisesti pienempi todennäköisyys sisältyä aineistoon verrattuna kokonaisaineistoon, koska kaikki perusjoukon havaintoyksiköt eivät valikoidu otokseen, vaan esimerkiksi joka 50:s.

Itsevalikoitunut otos: Etukäteen ei voi päätellä kuka tutkimukseen osallistuu, esim. netissä olevan linkin kautta. Tutkimuksilla tavoitellaan kuitenkin usein henkilöitä, joilla on kokemusta tutkittavasta asiasta. Näin tutkittavan asian laajuus vaikuttaa siihen, kuinka todennäköistä on, että voi päätellä jonkun henkilön kuuluvan aineistoon. Mieti esim. kokemuksia järjestötoiminnan johtamisesta vs. kokemuksia terveyskeskuspalvelusta.

Riippumatta perusjoukosta tai otantamenetelmästä on kuitenkin aina tärkeä tärkeää tutkia, mitä suoria tai epäsuoria tunnisteita aineistossa on ja käydä läpi aineisto harvinaisten tai ainutkertaisten havaintojen varalta.

Vastausprosentin suuruuteen on myös hyvä kiinnittää huomiota, sillä se kertoo havaintoyksikön todennäköisyydestä kuulua aineistoon. Erityisesti kokonaisaineistoissa tieto on merkittävä anonymiteetin arvioinnissa. Mitä korkeampi vastausprosentti, sitä todennäköisemmin havaintoyksikkö on aineistossa.

Tiedot aineiston muodostamistavasta eli otannasta- tai poiminnasta eivät saa paljastaa tutkimukseen osallistuneita. Tällaiseen voi olla riski, jos tutkija on kerännyt osallistujat lähipiiristään esim. lumipallomenetelmällä tai väestöllisesti pieneltä alueelta.

**2. Aineiston sisältö**

Voit kysyä:

**a.** Mitä suoria ja epäsuoria tunnisteita aineisto sisältää? Mitä aineiston tietoja yhdistelemällä henkilö saattaa olla tunnistettavissa?

Kartoita mitkä ovat ne tiedot, joiden perusteella henkilö voidaan mahdollisesti tunnistaa aineistoista. Tunnistamisen voi tehdä yksittäisen tiedon perusteella (suora tunniste) tai yhdistämällä tietoja. Ks. enemmän kohdasta [Milloin tieto on tunnisteellista](https://www.fsd.uta.fi/aineistonhallinta/fi/tunnisteellisuus-ja-anonymisointi.html#milloin-tieto-on-tunnisteellista). Voit jopa koittaa selvittää, voiko tietoja yhdistämällä henkilön tunnistaa. Huomaa, että tunnisteet eivät välttämättä sisälly vain kvantitatiivisen aineiston yksittäisiin muuttujiin tai kvalitatiivisen haastattelun alussa kerrottaviin henkilötietoihin. Kvantitatiivisissa aineistoissa niitä voi sisältyä satunnaisesti myös avomuuttujiin ja kvalitatiivisen haastattelun puheessa niitä voi tulla esiin useissa eri kohdissa.

Anonymisoinnissa tavallisesti ensimmäinen toimenpide on poistaa aineistosta suorat ja vahvat epäsuorat tunnisteet (ks. tunnistetaulukko). Suorien ja vahvojen epäsuorien tunnisteiden poistaminen riittää vain hyvin harvoin tekemään aineistosta anonyymin. Epäsuorien tunnisteiden tarkkuus ja määrä vaikuttavat anonymisointiin. Aineistoon sisältyvien henkilöiden tietoja tulee aina tarkastella suhteessa toisiinsa. Jo muutaman taustatiedon yhdistäminen voi tehdä henkilön tunnistettavaksi. Esimerkiksi sukupuoli, ikä, kunta ja tulot voivat paljastaa pienessä kunnassa suurituloiset henkilöt.

**b.** Sisältääkö aineisto kolmansiin henkilöihin liittyviä tietoja ja voiko niiden perusteella tunnistaa henkilöitä?

Kiinnitä huomio myös tietoihin, joita aineistossa tulee esiin kolmansista osapuolista. Joissain tilanteissa vastaajan henkilöllisyys voi paljastua kolmannen henkilön tietojen perusteella ja toisissa taasen kolmannen henkilön voi päätellä vastaajan tietojen perusteella. Yleisellä tasolla viitatut julkisuuden henkilöt voidaan jättää aineistoon. Jos on epävarma julkisesti tunnetun henkilön nimen jättämisestä aineistoon, voi arvioida, onko kerrottu tieto julkisesti tiedossa ja yhteiskunnallisesti niin merkittävä, että sen voi jättää aineistoon.

Ei me olla oltu mitenkään kovin uskonnollisia, vaikka tätini oli ensimmäisiä naispappeja, joita vihittiin Suomessa.

Tieto siitä, että täti oli ensimmäisiä naispappeja voi lisätä vastaajan tunnistamisriskiä, sillä ensimmäisistä naispapeista on julkisuudessa tietoa. Tosin ensimmäisessä pappisvihkimyksessä naisia oli 94, mikä on suhteellisen paljon. Riippuu aineistoon sisältyvistä tutkittavan muista taustatiedoista, tuleeko tieto poistaa vai ei.

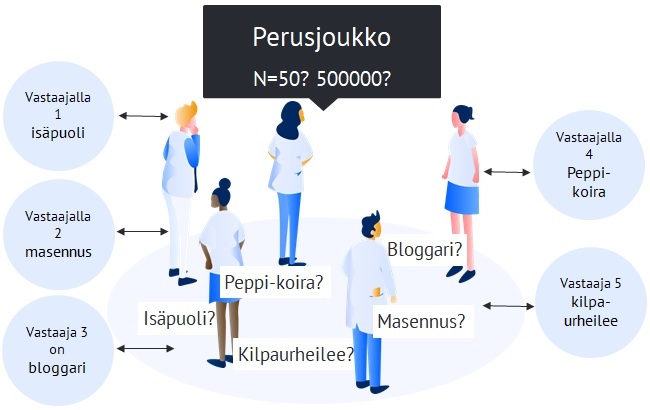
Kaupungilla kävellessäni törmäsin usein Satu Hassiin ja joskus hänen kanssaan vaihdoinkin pari sanaa.

Kansanedustajan voi jättää nimellä tekstiin, jos kansanedustajasta ei kerrota yksityisiä tietoja ja kaupungin mahdollinen paljastuminen ei uhkaa vastaajan anonymiteettiä. Kaupungin voitaisiin päätellä olevan julkisuudessa olevien tietojen perusteella Hassin asuinkaupunki Tampere tai työpaikan kaupunki Helsinki.

**c.** Sisältääkö aineisto harvinaisia tai ainutlaatuisia tietoja?

Havainnon harvinaisuus muodostuu yksittäisen tiedon tai kumuloituvien tietojen perusteella ja on aina riippuvainen kohdejoukosta. Tieto on harvinainen, jos sen esiintyvyys kohdejoukossa on pieni. Harvinainen tieto tulee anonymisoida etenkin, jos tiedon voi saada selville ulkopuolelta saatavia tietoja hyväksi käyttäen.

Seuraavan kuvion avulla havainnollisestaan vastaajien ja kohdejoukon välistä tietojen arviointia. Vastaajista kertyy aineistoissa monenlaista toisistaan poikkeavaa tietoa, kuten että vastaajalla on isäpuoli, Peppi-niminen koira, hän sairastaa masennusta tai että hän on bloggari tai kilpaurheilija. Vastaajista saatavien tietojen kohdalla tulee arvioida, onko tieto yleinen vai harvinainen kohdejoukossa.



Jos kohdejoukko on 2000 vastaajan otos Suomen väestöstä, tiedot isäpuolesta ja masennuksesta ovat yleisiä kohdejoukossa eli koko väestössä. Harvinaisempia tietoja ovat Peppi-nimisen koiran omistaminen, kilpaurheilu ja bloggaus. Riippuen muista henkilöistä saatavista tiedoista, edellä mainittuja tietoja tulee anonymisoida. Jos kohdejoukko on tarkemmin rajattu joukko henkilöitä kuten pienen kyläkoulun oppilaat, tieto isäpuolesta, puhumattakaan kilpaurheilusta, bloggaamisesta tai Peppi-koirasta tekevät henkilön hyvin todennäköisesti tunnistettavaksi.

Masennus on esimerkki tiedosta, joka ei yleensä tarvitse anonymisointia, koska tiedon luonne on näkymätön. Sitä ei ulkopuolelta voi välttämättä nähdä ihmisestä, eikä tietoa useinkaan jaeta kuin vain läheisimmille ihmisille. Tieto masennuksesta voi olla silti tunnisteellinen esimerkiksi työpaikkakyselyissä, jos henkilö on joutunut olemaan masennuksen vuoksi pitkään sairaslomalla.

Harvinainen tieto ei ole siis automaattisesti tunnisteellista, sillä harvinainen tieto ei ole aina julkisesti saatavilla olevaa tietoa.

Väestötason aineistossa selviää, että yhdellä vastaajalla on katoptrofobia eli peiliin kohdistuva pelko.

Tieto on harvinainen, mutta ei välttämättä tunnisteellinen, sillä tiedon yhdistäminen tiettyyn henkilöön on vaikeaa, koska kyseistä tietoa ei ole saatavilla julkisesti.

Väestötason tutkimuksissa harvinaisia tietoja voivat olla esimerkiksi henkilön harvinainen ammatti tai asema esimerkiksi organisaation tai yrityksen johtajana tai korkeassa asemassa olevana poliitikkona. Harvinaisia tietoja voivat olla myös, suuret tulot tai omaisuus, sairaudet, kilpaurheilu tietyssä lajissa, harrastus tai osallisuus jossain tapahtumassa, josta on ollut tietoa mediassa.

Väestötutkimuksissa anonymiteettiä voi rajata tehokkaasti aluetietoja muokkaamalla.

Golfin harrastaja väestötason tutkimuksessa ei ole harvinainen.

Golfin harrastaminen Pirkanmaan alueen rehtoreiden keskuudessa on suurella todennäköisyydellä jo harvinainen tieto. Sitä saattaa harrastaa esimerkiksi vain pari rehtoria. Golf-harrastus saattaa olla mainittu myös esimerkiksi paikallislehden rehtorin haastattelussa.

Joissain tutkimuksissa itse tutkittava ilmiö on harvinainen ja tällöin tulee huolehtia kohdejoukon riittävästä suuruudesta. Esimerkiksi tutkittaessa talvilajien huippu-urheilijoita alueen kattavuudella on suuri merkitys anonymiteetille, sillä urheilijoiden määrä vaihtelee yksittäisen kunnan, maakunnan, Suomen tai koko maailman mittakaavassa. Sitä anonyymimmän aineiston saa, mitä laajempi harvinaisen ilmiön tutkimuksen maantieteellinen alue on.

Käytännössä jokaisen vastaajan kyselytutkimuksen arvot tai laadullinen aineisto voivat tuottaa ainutlaatuisia tietokokonaisuuksia, jollaisia ei ole missään muualla maailmassa. Jos tiedot olisivat saatavilla ulkopuolelta, henkilön voisi tunnistaa. Kyselytutkimusten mielipiteitä ja asenteita koskevia väittämiä on kuitenkin hankala saada ulkopuolelta selville, sillä ihmiset myös unohtavat kyselyihin valitsemansa vastausarvot. Ihmiset myös saattavat muistaa ja kertoa tutkimushaastattelussa selittämänsä tapahtuman vuoden kuluttua toisella tavalla. Tärkeintä onkin arvioida, voiko henkilön tunnistaa tietoja yhdistämällä tai yhdistämällä niitä ulkopuolella oleviin tietoihin.

**d.** Ovatko aineiston tiedot sensitiivisiä?

**Tiedot ovat sensitiivisiä, mikäli ne käsittelevät tietosuojalainsäädännön mukaisiin erityisiin henkilötietoryhmiin lukeutuvia tietoja, joista ilmenee etninen alkuperä, poliittisia mielipiteitä, uskonnollinen tai filosofinen vakaumus tai ammattiliiton jäsenyys, geneettiset tiedot, henkilön tunnistamista varten käsitellyt biometriset tiedot, terveyttä koskevat tiedot ja seksuaalista käyttäytymistä ja suuntautumista koskevat tiedot. Myös muut tiedot voivat olla luonteeltaan sensitiivisiä. Arkaluonteisuutta voi mitata esimerkiksi sillä, onko asia yhteiskunnassa tabu tai kuinka paljon asian paljastumisesta voi koitua haittaa henkilölle, organisaatiolle tai muulle havaintoyksikölle.**

Esimerkkejä muista arkaluonteisista asioista ovat: rikostapahtumat kuten perheväkivaltakuvaukset, muihin henkilöihin kohdistuvat arvostelut, yksityiskohtaiset kuvaukset kolmansien henkilöiden elämänkulusta tai yrityssalaisuuksien piiriin kuuluvat asiat.

**3. Aineiston ikä**

Aineiston ikä vaikuttaa anonymisointitarpeeseen. Mitä vanhempi aineisto on, sitä hankalampaa henkilöiden tunnistaminen on, koska tiedot muuttuvat aikojen kuluessa. Yli 100-vuotta vanhoja tai kuolleiden henkilöiden tietoja ei tarvitse enää suojella.

**4. Vastaajista muualta saatavat tiedot**

Anonymisoinnin onnistumiseksi aineiston sisältämiä tietoja tulee tarkastella myös suhteessa muualta saatavissa oleviin tietoihin. Aineisto tulee muokata niin, ettei muualta saatavilla olevien tietojen perusteella voi tunnistaa yksittäistä henkilöä.

Oman aineiston tietoja kannattaa pohtia suhteessa neljään eri tietotyyppiin (Elliot ym. 2016):

1. samasta kohdejoukosta muulta löytyviin tietoihin ja tutkimusaineistoihin
2. julkisesti saatavilla oleviin tietoihin (esim. julkiset rekisterit ja sosiaalinen media)
3. ihmisten paikallistietämykseen (miltä asuinpaikat näyttävät ja mitä alueella tapahtuu)
4. henkilökohtaisiin tietoihin muista ihmisistä (mitä tiedän esim. naapureistani).

Mitä todennäköisemmin tiedot ovat saatavilla tai niistä tiedetään yleisesti, sitä enemmän aineiston sisältämiä tietoja tulee rajoittaa. Seuraavassa esitetään esimerkkejä erilaisista aineistosta ja niihin yhdistettävistä ulkopuolisista tietolähteistä:

a. Aineistossa käsitellään suomalaisten urapolkuja. Ulkopuolisia tietoja suomalaisten urapoluista on saatavilla internetistä mm. Linkedinistä, sosiaalisesta mediasta kuten Facebookista ja työpaikan nettisivujen henkilöstötiedoista.

b. Aineistossa käsitellään suomalaisten päivän ateriarytmiä. Ulkopuolisia tietoja ei ole helposti saatavilla, vaikka ateriarytmin tutkiminen voi tuottaa hyvin yksityiskohtaistakin tietoa henkilön arjesta. Olennaista on kysyä, miten muiden ihmisten ateriarytmiin liittyviä tietoja voi saada selville. Ei kovinkaan helposti tai ne muistuttavat toinen toistaan.

c. Aineistossa käsitellään suomalaisten ja tansanialaisten naapurussuhteita. Tässä tulee selvittää, miten suomalaisessa ja tansanialaisessa kulttuurissa tiedetään naapureiden asioista. Suomessa vuorovaikutus naapureiden kanssa voi olla hyvinkin vähäistä, mutta Tansaniassa voi olla toisin. Näin anonymisoitavaa on todennäköisesti enemmän Tansanian aineistossa.

Jos aineistosta on julkaistu jo raportteja tai julkaisuja, ota huomioon, kuinka yksityiskohtaisesti aineiston tiedot on esitetty niissä.

Eräässä määrällisessä aineistossa anonymisoinnissa on päätetty muokata kuntamuuttujaa niin, että muuttujasta poistetaan kunnan nimi ja se jää aineistoon pelkkänä numeroarvona (1, 2, 3...) esim. monitasoanalyysejä varten. Sitten huomataan, että aineistosta aikaisemmin tehdyssä julkaisussa kerrotaan vastaajien määrät eri kunnissa. Anonymisointitoimenpide on näin epäonnistunut, sillä kuntien nimet ovat palautettavissa vastaajien määrän avulla takaisin.

Aineiston ulkopuolisilla tiedoilla on siis suuri merkitys anonymisoinnissa. Eräs valaiseva erimerkki ulkopuolisten tietojen yhdistettävyydestä on Latanya Sweeneyn (2000) tutkimus, jossa havaittiin, että yhdysvaltalaisista 87 % on tunnistettavissa syntymäajan, sukupuolen ja viisinumeroisen postinumeron (ZIP-koodin) perusteella. Tunnistaminen perustui äänestyslistoihin, jotka Sweeney tilasi tutkimustaan varten. Listat sisältävät henkilö- ja aluetietoja äänioikeutetuista. Hieman yli puolet Yhdysvaltojen populaatiosta eli 53 % on puolestaan tunnistettavissa vain sukupuolen, syntymäajan ja paikan, kuten kaupungin tai kylän, avulla (em.).

**5. Käytettävyys vs. anonymiteetti**

Anonymisointi poistaa aina tietoja aineistosta. Mitä enemmän anonymisointia, sitä enemmän aineiston käytettävyys ja tulosten tarkkuus kärsivät. Ihannetapauksessa aineistoon tehdään mahdollisimman pieniä muutoksia välttäen tutkimuksellisesti tärkeimpien muuttujien muokkaamista. Usein se on helpommin sanottu kuin tehty.

Onnistuneessa anonymisoinnissa aineiston käsittelijän tunnistaa tutkimuksen kannalta erittäin merkittävät ja vähemmän merkittävät tiedot nykyisen ja tulevaisuuden tutkimuksen kannalta. Tunnisteiden poistoa pyritään tekemään vähemmän merkityksellisiin tietoihin. Joskus esimerkiksi määrällisissä aineistoissa avomuuttujien pudottaminen poistaa useat harvinaiset ja ainutlaatuiset tiedot. Numeeriset muuttujat ovat määrällisessä tutkimuksessa usein myös helpommin käytettäviä kuin avovastaukset. Usein anonymisointia joutuu tekemään myös merkittäville muuttujille, kuten iälle anonymiteetin saavuttamiseksi.

Jos aineistoon haluaa jättää asuinkunnan, täytyy anonymisoinnin ratkaisut toteuttaa henkilöä koskevien muiden taustatietojen osalta. Siten ammatti, työpaikka, koulutus, ikä, jne. luokitellaan riittävän karkealle tasolle. Jos taas on sisällöllisesti tärkeää jättää aineistoon tieto tutkittavien ammatista ja iästä, tulee tutkittavia koskevat aluetiedot karkeistaa (kuntatiedon sijaan suuralue ja/tai kuntatyyppi) ja arvioida myös muiden taustatietojen muokkaustarve.

**Anonymisoinnin varmistus**

Elliotin ym. (2016 ) sanoin "anonymisointi ei ole eksaktia tieteen harjoittamista", joten riittävän anonymisoinnin arviointi voi olla joskus ongelmallista. Anonymisoinnin menetelmän valintaa ja toteutuksen onnistumista voi kuitenkin arvioida tehokkaasti seuraavien kysymysten avulla (sovellettu EU:n WP 29 lausuntoa 05/2014). Jos vastaus kahteen ensimmäiseen on kielteinen ja viimeisessä todennäköisyys päättelyyn hyvin pieni, aineiston anonymiteetti on hyvällä mallilla.

1. Havainnon erottaminen joukosta: Voiko anonymisoinnin jälkeen henkilö olla edelleen tunnistettavissa aineistosta?
2. Yhdistettävyys: Voiko vastaajien tiedot yhdistää toiseen aineistoon tai ulkopuoliseen tietoon, ja mahdollisesti sitä kautta tehdä henkilöitä tunnistettaviksi?
3. Päättely: Voiko tietojen päätellä koskevan tiettyä henkilöä? Onko muokattujen tai poistettujen tietojen alkuperäiset arvot pääteltävissä?

**Muista!** Koska avoimesti saatavilla oleva tieto lisääntyy koko ajan, on tärkeää tarkistaa säännöllisesti, onko kertaalleen anonymisoitu aineisto edelleen anonyymi (ns. jäännösriskin arviointi).

Lomakkeen yläreuna

Lomakkeen yläreuna