

Kvalitatiivisen aineiston anonymisointikeinot

Katja Fält

5.4.2017

Tampereen yliopisto

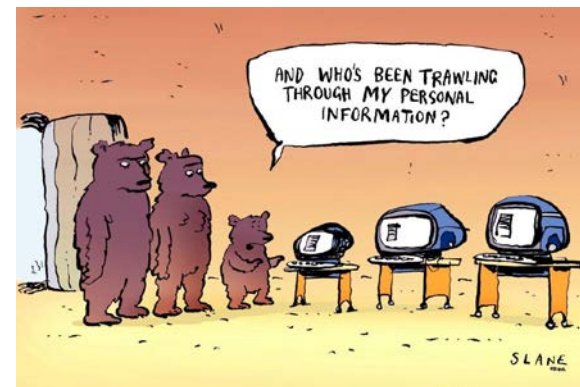


TIETOARKISTO



Henkilötieto

- ❖ Henkilötieto: kaikenlaiset luonnollista henkilöä tai hänen ominaisuuksiaan tai elinolosuhteitaan kuvaavat merkinnät, joista henkilö, hänen perheensä tai hänen kanssaan yhteisessä taloudessa elävät voidaan tunnistaa (Henkilötietolaki 3.1.§)
 - Kaikki luonnolliseen henkilöön liittyvät tiedot voivat olla henkilötietoja
 - Tiedon luonnetta ei ole rajattu (voivat olla myös subjektiivisia mielipiteitä, väitteitä tai arvoasetelmia, ei edellytetä totuutta)
- ❖ Tutkimusaineistoissa henkilötiedot voivat koskea sekä tutkittavia että kolmansia osapuolia





Tunnisteellinen tieto

- ❖ Tieto tunnisteellista, jos sen perusteella voidaan tunnistaa yksittäinen henkilö
 - Henkilö voidaan tunnistaa yhden tai useamman tunnusomaisen fyysisen, psyykkisen, taloudellisen, kulttuurisen tai sosiaalisen tekijän perusteella
 - Suorat ja epäsuorat tunnisteet
- ❖ Henkilötietolain mukaan myös pseudonyymit aineistot ovat tunnisteellisia



Tunnisteellinen tieto

Suorat tunnisteet

- ❖ Esimerkiksi koko nimi, henkilötunnus, henkilönimen mukainen sähköpostiosoite, biometriset tunnisteet
 - **Vahvat epäsuorat tunnisteet:**
 - Yksittäisiä tietoja, joiden perusteella henkilön voi kohtuullisen helposti tunnistaa
 - esim. osoite, puhelinnumero, auton rekisteri, harvinainen ammattinimeke, harvinainen sairaus

Epäsuorat tunnisteet

- ❖ Sukupuoli, ikä, asuinkunta, ammattinimike
- ❖ Voivat mahdollistaa tunnistamisen yhdistettyinä muihin tietoihin



Anonyymi tieto

- ❖ Tunnusomaiset piirteet (esim. epäsuorat tunnisteet yhdistettynä) koskevat samanlaisina useampaa henkilöä

”Pirkko”

Syntymäaika: 1976

Asuinpaikka: Tampere

Ammatti: Tieteenala-asiantuntija

Työpaikka: Tampereen yliopisto

Koulutus: Filosofian tohtori

”Pirkko”

1976

Pirkanmaa

Erityisasiantuntija

Yliopisto

Ylempi korkeakoulututkinto





Anonyymi tieto

- ❖ Tunnusomaiset piirteet (esim. epäsuorat tunnisteet yhdistettynä) koskevat samanlaisina useampaa henkilöä
- ❖ Katsotaan että henkilöä ei voida tunnistaa huomioiden kohtuullisesti toteutettavissa olevat toimenpiteet
 - Tietosuoja valtuutettu: edes tutkittava itse ei saisi tunnistaa itseään aineistosta
 - käytännössä näin anonyymia aineistoa on vaikea tuottaa
- ❖ Anonymisoinnin tulee olla peruuttamaton, jotta voidaan puhua anonyymista aineistosta





Women Wearing Paper Bags on Their Heads, 1967
© Bettmann/CORBIS

Tunnistetietojen käsittely

- ❖ Tunnisteellisia tietoja voi käyttää tieteellisen tutkimukseen, kun se on tarkoituksenmukaista, suunniteltua ja perusteltua
- ❖ Tunnisteet ja niiden säilyttämisaika määritettävä rekisteriselosteessa
- ❖ Käsittelyssä korostuu suunnitelmallisuus
- ❖ Tarpeettomat tunnisteet poistetaan aina kun se on mahdollista
- ❖ Tunnistetietojen käsittely aina aineistokohtaista
- ❖ Tunnisteellisten tietojen jättäminen tutkimusteksteihin voi olla joissain tapauksissa perusteltua (mm. historian tutkimus, asiantuntijahaastattelut) → sovittava



Anonymisoinnin lähtökohdat

- ❖ Tapauskohtaisuus
- ❖ Suorien ja vahvojen epäsuorien tunnisteiden poisto harvoin tekee aineistosta anonyymin
- ❖ Lisäksi huomioitava epäsuorien tunnisteiden poisto ja muokkaaminen
 - Mitä enemmän tunnisteita, sitä huolellisemmin anonymisointi tulee tehdä
- ❖ Yksittäisiä henkilöitä ei saa tunnistaa aineistosta muualta saatavien tietojen perusteella
 - Myös kolmannet osapuolet
- ❖ Tunnisteet anonymisoitava myös mahdollisista taustamateriaaleista



Laadullisen aineiston anonymisointikeinoja

1. Poistaminen
2. Pseudonymisointi
3. Kategorisointi
4. Tunnistetietojen vaihtaminen





Tunnistetietojen poistaminen

- ❖ Suorat tunnisteet poistetaan
- ❖ Tekstiaineistot: tunnisteita sisältävät taustamateriaalit hävitetään (tutkittavien yhteys- ja taustatiedot)
- ❖ Poistetaan myös: tiedostojen piilotetut tekijä-, kuva- ja paikkatiedot

- ❖ Yksilöivät, arkaluontoiset tiedot voidaan poistaa
 - Perusteltua kun
 1. tiedot paljastuvat sattumalta
 2. Tiedot eivät ole tutkimuksen kohteena
 3. Tieto muodostaa tunnistamisriskin



Pseudonymisointi

- ❖ Henkilönimet vaihdetaan peitenimiksi eli pseudonyymeiksi
 - Matti -> Pekka, Anna -> Liisa
 - Suunniteltava johdonmukaisesti
 - Parempi ratkaisu kuin nimien poistaminen kokonaan, aineiston ymmärrettävyys säilyy
- ❖ Julkisesti tunnettuihin henkilöihin voidaan viitata nimellä
 - Pseudonymisointi (tai kategorisointi) voidaan tehdä, mikäli aineisto sisältää henkilön yksityiselämään liittyviä arkaluonteisia tietoja





Kategorisointi

- ❖ Taustatietojen kategorisointi eli karkeistaminen yleisemmälle tasolle
 - Tärkeää kontekstitietoa, joiden poistaminen voi heikentää aineiston tulkintaa
 - Henkilönimien karkeistaminen (jos esiintyy aineistossa vain kerran/muutaman kerran, eivät ole keskeisiä aineiston sisällön ymmärrettävyyden kannalta
 - Marjatta → [sisko], [naapuri]
 - Yksilöivät työpaikat, ammattinimikkeet, toimipaikat, oppilaitokset, paikkakunnat, jne.
 - Tampere → [kaupunki eteläisessä Suomessa]
 - Ravintola Tillikka → [ravintola]
 - Koskikeskus → [ostoskeskus]
 - Hankeasiantuntija → Tutkimushallinnon asiantuntija



Tunnistetietojen vaihtaminen

- ❖ Esimerkiksi tarkka syntymäaika (joka tunnisteena tulisi poistaa) voi kuitenkin joskus olla merkityksellinen aineiston sisällön ymmärtämiselle
→ tietoja on mahdollista muuttaa

Syntymäaika 28.4.1976 → 28.4.1978





Tunnistetaulukko

Tunnistetieto	Suora tunniste	Vahva epäsuora tunniste	Epäsuora tunniste	Anonymisointitekniikka
Henkilötunnus	x			Poista
Koko nimi	x			Poista/Muuta
Sähköpostiosoite	x	x		Poista
Puhelinnumero		x		Poista
Postinumero			x	Poista/Luokittele
Kaupunginosa			x	Luokittele
Asuinkunta			x	Luokittele
Maakunta			x	(Luokittele)
Suuralue			x	
Tilastollinen kuntaryhmä			x	
Puhetallenne	x			Poista
Videotallenne henkilö(i)stä	x			Poista
Valokuva henkilöstä	x			Poista
Syntymäaika		x		Luokittele
Ikä			x	Luokittele
Sukupuoli			x	
Siviilisääty			x	
Perheen koostumus			x	(Luokittele)
Ammatti		(x)	x	Luokittele

Toimiala			x	
Työmarkkina-asema			x	
Koulutus			x	Luokittele
Koulutusala			x	
Äidinkieli			x	Luokittele
Kansallisuus			x	(Luokittele)
Työpaikka		(x)	x	Luokittele
Auton rekisteri		x		Poista
Tutkittavan julkaisun/teoksen nimike		x		Luokittele
Verkkosivun osoite		(x)	x	Poista
Opiskelijatunnus		x		Poista
Vakuutusnumero		x		Poista
Tilinumero		x		Poista
Tietokoneen IP-osoite		x		Poista
Terveyttä koskevat tiedot *		(x)	x	Luokittele/Poista
Etninen alkuperä *		(x)	x	Luokittele/Poista
Rikos tai saatu rangaistus *			x	Luokittele/Poista
Ammattiliiton jäsenyys *			x	Luokittele
Poliittinen tai uskonnollinen vakaumus *			x	Luokittele
Muu luottamustoimi tai jäsenyys		(x)	x	Luokittele/Poista
Sosiaalihuollon tarve *			x	Luokittele/Poista
Sosiaalihuollon tukitoimet ja etuudet *			x	Luokittele/Poista
Seksuaalinen suuntautuminen *			x	Poista

<http://www.fsd.uta.fi/aineistonhallinta/fi/tunnisteellisuus-ja-anonymisointi.html#tunnistetaulukko>



Käytännön vinkit

- ❖ Tee kirjallinen anonymisointisuunnitelma
- ❖ Tarkista aineistotiedostoihin liittyvät taustamateriaalit
- ❖ Muista tarkistaa, ettei aineistoon jää tunnistetietoja kolmansista henkilöistä
- ❖ Tee muutokset yksi kerrallaan
- ❖ Tarkista, esiintykö aineistossa sama henkilö eri nimillä
- ❖ Helpota tekstitiedoston anonymisoinnin suunnittelua kirjaamalla litteraattiin erikoismerkki jokaisen henkilö- ja erisnimen eteen
- ❖ Huolellisuus palkitaan



KIITOS!

[asiakaspalvelu.fsd @ uta.fi](mailto:asiakaspalvelu.fsd@uta.fi)