



TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE

1

Tutkimusaineistojen anonymisoinnin työpaja

Avoim TKIO -verkosto 16.3.2026

SANNI TUJUNEN & EMILIA HAKKOLA
TIETOARKISTO



Työpajan ohjelma

- > Anonymisoinnista yleensä
- > Kvalitatiivisen aineiston anonymisointi ja ryhmiin jako
- > Ryhmätyö kvalitatiivisen aineiston anonymisoinnista n. 13 min
- > Tehtävän purku
- > Kvantitatiivisen aineiston anonymisointi ja ryhmiin jako
- > Ryhmätyö kvantitatiivisen aineiston anonymisoinnista n. 13 min
- > Tehtävän purku
- > Loppukeskustelu

Anonymisoinnista yleensä



TIETOARKISTO





Vääriä oletuksia anonymisoinnin tarpeesta

- > Tunnistamisen tarkoitukset ja seuraukset
- > Ei-sensitiivistä sisältöä
- > Vain muutama henkilö tunnistettavissa



Anonyymi aineisto

- > Anonyymistä aineistosta ei voi tunnistaa henkilöitä **suoraan tai välillisesti** kohtuullisen todennäköisin keinoin.
 - Suoraan tunnistaminen voi tapahtua esimerkiksi henkilötunnuksen tai valokuvan perusteella.
 - Välillinen tunnistaminen voi onnistua aineiston sisältämien epäsuorien tunnisteiden tai niiden ja muualta saatavien tietojen avulla.
- > Kohtuullisen todennäköisiä tunnistamisen keinoja arvioitaessa voidaan pohtia, **mitä tietoja ihmisillä on käytettävissä aineiston lisäksi**.
 - Suurin osa ihmisistä pääsee julkisiin tietolähteisiin. Näitä ovat erilaiset viralliset lähteet, hakukoneella löydetty verkkosisältö sekä sosiaalinen media.
- > Anonymisoinnin tulisi olla peruuttamaton
- > Pseudonyymi aineisto tulkitaan henkilötietoja sisältäväksi

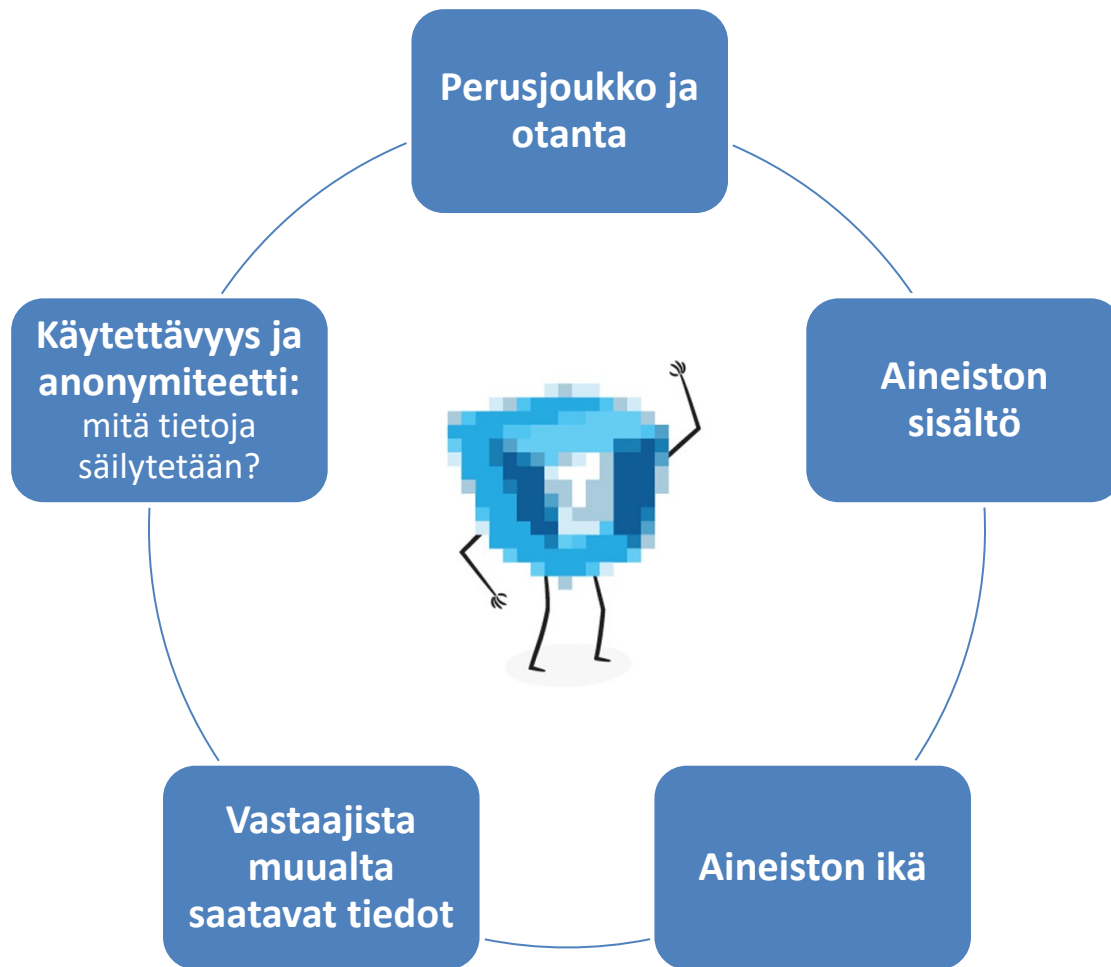


Minimointi

- > Anonymisointi helpottuu, kun alkujaan välttää turhien, yksityiskohtaisten tai tutkimuksen kannalta merkityksettömien tietojen keräämistä
- > Suunnittele huolella:
 - Mitä taustatietoja henkilöistä tarvitaan tutkimukseen?
 - Älä käytä avokysymyksiä, jos vastaukset voi luokitella
- > <https://www.fsd.tuni.fi/fi/palvelut/aineistohallinta/tunnisteellisuus-ja-anonymisointi/#minimointi-eli-miten-kerata-aineisto-niin-ettei-se-sisalla-turhia-tunnisteita>



Mitä asioita tulee ottaa huomioon suunniteltaessa anonymisointia?





Anonymisointitapoja

Yleistävät menetelmät:

- luokittelu (esim. iän luokittelu 5 vuoden välein). Luokittelussa pyritään jättämään yksityiskohtaisin luokitus, esim. ei luokitella ikää 10 vuoden välein, jos ei ole pakko.
- poistaminen: tietojen/muuttujan/havaintoyksiköiden
- kategorisointi esim. [taajama], [metallialan yritys], [yläkoulu], [kauppakeskus]

Henkilönimien korvaaminen peitenimellä: Katri → [Saara].

Sekoittavat menetelmät:

- esim. luokitellaan ikä +-2 vuotta, muutetaan tapahtuman päivämäärää tai vaihdetaan vastaajien tietoja keskenään. Käytettävä harkiten!

Anonymisoinnissa poistetaan aina suorat tunnisteet!

Poistettava tieto ei saa olla pääteltävissä!

Kvalitatiivisen aineiston anonymisointi



TIETOARKISTO





Poista/karkeista suorat tunnisteet ja vahvat epäsuorat tunnisteet

- Tutkittavien nimet, äänitiedostot, videotiedostot
- Yhteystiedot, täsmälliset syntymäajat, IP-osoitteet ja muut tunnisteet
- Kirjoitusaineistoista piilotetut tekijätiedot (Ms Office ohjeet löytyy)
- Kuva-aineistojen tekijätiedot (Exif-tiedot)
- Anonymisoinnin lopuksi suostumuslomakkeet (jos käytetty)

Vahvat epäsuorat tunnisteet

- Harvinaiset nimikkeet, asemat
- Viittaukset teoksiin ja uutiskynnyksen ylittäneisiin tapahtumiin



Anonymisoinnin keinoja (Osa 1)

> Taustatietojen luokittelu

- Henna Juusola, tietoasiantuntija Tietoarkistossa, 31 v, asuu Lempäälässä →
- [nainen] tai [Noora], asiantuntija yliopistossa, ikäryhmä: 30-34-vuotias, asuinpaikka: kaupunkimainen kunta, Etelä-Suomi

> Henkilönimien vaihtaminen keksityiksi nimiksi

- Voi halutessa jättää tiedon sukupuolesta Henna -> [Noora]
 - Ylen nimikone <https://yle.fi/a/3-9418798>
- Voi halutessa jättää kulttuurista tietoa (Alev -> [Damla])
- Huom! Merkkaa muutetut kohdat, jotta jatkokäyttäjä tietää, mitä tietoa on anonymisoitu



Anonymisoinnin keinoja (osa 2)

- > Erisnimien kategorisointi
 - Henna → [nainen, naapuri], [nainen, työtoveri], [vaimo], [sisko], [täti]
 - TAMK → [ammattikorkeakoulu], [korkeakoulu]
 - Tammelan koulu → [alakoulu]
 - Tavara-asema → [konsertti- ja tapahtumatalo]
 - Pulteri → [ravintola]
- > Aineisto-osuuksien poistaminen
 - [Poistettu kolmatta henkilöä koskevaa tunnistesteellista tietoa]
- > Aluetietojen luokittelu, Tilastokeskukselta löytyy luokitusavaimia
 - Seinäjoki → [kaupunkimainen kunta, Länsi-Suomi]
 - Alavus → [taajaan asuttu kunta]



Aineistoesimerkki ryhmätyöhön

<https://www.fsd.tuni.fi/files/kvalitatiivinen-aineistoesimerkki.docx>



Yksi malli anonymisoinnista

- > Haastateltava: [Heidi]
- > Ammattinimike: [-] asiantuntija
- > Työpaikka: [-] yliopisto [yksikkö]
- > Ikäryhmä: [35-39 v]
- > Asuinpaikka: [Kaupunki, Etelä-Suomi]
- > Perhe: puoliso ja kaksi alaikäistä lasta
- > Haastattelun päivämäärä: 16.12.2023
- > Tutkija: Maija Kiireinen
- >
- > T: Jos pohdit alkuperäistä opiskelualaasi, miten arvioit perhetaustasi vaikuttaneen sen valintaan?
- > H: Ärsyttää tuo koko kysymys. Tai jos nyt mietin niin ehkä onkin vaikuttanut jossain määrin. Mun isähän on tuossa [-] sairaalassa ollut pitkään ja edelleen [johtavassa asemassa] ja ei kyllä yhtään toivonut, että lähtisin lukemaan historiaa. Ei tosin määrännyt hakemaan lääkikseen, mutta on kyllä ollut selvä näkemys hällä, että vain isoa palkkaa tuottavat alat on järkeviä. Se näkyy kyllä hänen asenteissaan ja siinä, miten esimerkiksi työttömistä puhuu.



- > T: Miten sitten äiti?

- > H: Ei oo sisaria. Äiti on apteekkari ja siitä voit jo päätellä, ettei oo ihan köyhä tää mun lapsuuden perhe. Hän ei ehkä niinkään korostanut sitä rahapuolta, mutta jotenkin menestys ja asema on hälle merkittäviä arvoja. On ollut nyt kans [\[yhdistyksen\]](#) puheenjohtaja kolme vuotta ja se on kyllä vaikuttanut jotenkin aivan liikaa hänen olemukseensa. Keekoilee kotonaanki ku joku Karita Mattila.

- > T: Mutta yrittikö äiti jotenkin vaikuttaa sun alan valintaan tai miten suhtautui siihen?

- > H: No pettymys oli kasvoilta nähtävissä, kun kerroin, että historia kiinnostaa ja sitä haluan opiskella.

- > T: hmm....jos ajattelet valintaasi, oliko se osin jotain kapinaa tai...?

- > H: No saattoi olla vähän sitäkin, etten halunnut just jatkaa ja olla mukana heidän kuvitelmissaan. Ehkä oli jotain kapinaa. Ehkä on vieläkin, kun mehän asutaan vuokralla. Asutaan siis [\[kaupunginosassa\]](#) [\[kahden kadun\]](#) kulmauksessa. Satavuotias ihana punainen puutalo. Vuokraisäntä on tehnyt pakolliset remontit, mutta ei mun vanhemmat arvosta semmosta että asuu talossa, jossa olisi korjattavaa ja varsinkaan sitä vuokralla asumista eivät voi sietää.

- > T: On kai teillä kuitenkin jotain yhteistä niinkun sun vanhempien kanssa?

- > H: Sanosin, ettei oo paljonkaan. Esimerkiksi kun heille viimeksi selitin että oon lukemassa nyt koko Anne Vuori-Kemilän tuotantoo ja he kysyy, että mistä se kirjottaa. Kun sitten alan kertoo, että köyhistä pohjoisen ihmisistä 1900-luvun alkupuolella, niin heti alkaa heillä silmät pyöriä rasittuneina.

Kvantitatiivisen aineiston anonymisointi



TIETOARKISTO



Käytännön anonymisointivinkkejä kvantitaiaineistoille 1/2

- Suunnittele anonymisointi ottaen huomioon kaikki aineiston muuttujat
 - ei vain nk. taustamuuttujat. Silmäile myös avomuuttujien vastausten sisältöjä.
- Tee anonymisointi alkuperäisen datan kopioon
- Anonymisoi ensin numeeriset muuttujat ja viimeisenä avomuuttujat
 - numeeristen muuttujien muokkaus vaikuttaa siihen, miten avomuuttujia anonymisoidaan
 - avomuuttujat eivät saa sisältää numeerisia muuttujia tarkempia tietoja esim. iästä
- Anonymisointipäätöksissä ota huomioon muuttujien tilastollinen käyttökelpoisuus ja tieteellinen kiinnostavuus
 - Esim. kuinka paljon avomuuttujia käytetään kvantitutkimuksessa?



Käytännön anonymisointivinkkejä kvantitieteille 2/2

- Ole huolellinen harvinaisten arvojen poistamisessa/muokkaamisessa
 - Esim. yhtä harvinaista luokkaa ei voi laittaa yksistään SYSMISIin, sillä dokumentoinnin perusteella tieto on palautettavissa. SYSMIS-arvoissa tulee olla alun perin havaintoja tai tietoa pitää poistaa useammasta luokasta.
- Kirjaa tehdyt anonymisointitoimenpiteet ylös
 - esim. syntaksiin tai anonymisointisuunnitelmaan, jotta tehdyt toimenpiteet on dokumentoitu
- Käytä avovastausten anonymisointiin jotain merkintää, esimerkiksi [hakasulkuja], jotta tiedetään mitä on muutettu ja mitä ei
 - Joskus on perusteltua tehdä anonymisointia ilman merkintää, jos merkinnästä voi päätellä poistetun asian



Aineisto ryhmätyöhön:

<https://www.fsd.tuni.fi/files/kvantitatiivinen-aineistoesimerkki-lyhennetty.docx>



Kvantitatiivinen aineistoesimerkki: anonymisointiehdotus 1/3

Vaihtoehto 1: säilytetään kuntamuuttuja

LUOKITELLAAN:

- **Ikä:** luokitellaan 3 vuoden välein
- **Talouden koko:** TOP-koodataan esim. 6 tai yli

Sukupuoli ja kunta jätetään sellaisenaan aineistoon

TAI **Kunta:** pienimmät kunnat Puumala ja Salla yhdistetään

POISTETAAN:

- Synnyinmaa

TAI Voidaan tehdä myös niin, että **synnyinmaata ei poisteta vaan** luokitellaan 1 suomi, 2 muu, **JOS** pienimmissä kunnissa ei ole tässä aineistossa muualla kuin Suomessa syntyneitä.



Kvantitatiivinen aineistoesimerkki: anonymisointiehdotus 2/3

Vaihtoehto 2: Luokitellaan kunta → maakunta?

MUTTA koska aineistossa on mukana myös kunnan asukasmäärä -muuttuja, osan kunnista kohdilla tieto on pääteltävissä.

JOKO:

1. Tehdään samat toimenpiteet kuin vaihtoehdossa 1. TAI
2. Poistetaan kunnan asukasmäärä, jolloin
 - **Ikä** saa jäädä sellaisenaan
 - **Talouden koko** TOP-koodataan esim. 6 tai yli
 - **Synnyinmaa** luokitellaan 1 Suomi, 2 Muu



Kvantitatiivinen aineistoesimerkki: anonymisointiehdotus 3/3

Vaihtoehto 3: Ei kuntamuuttujaa!

Kaikki muut tiedot voidaan jättää sellaisenaan.

Kiitos!

Kysymyksiä?

Ota yhteyttä: asiakaspalvelu.fsd@tuni.fi



Tietoarkisto



TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE