



TIETOARKISTO  
FINNISH SOCIAL SCIENCE  
DATA ARCHIVE

# Anonymisointi

**Avoim TKIO 18.3.2021**

LUENNOITSIJAT  
EMILIA HAKKOLA & ARJA KUULA-LUUMI  
TIETOARKISTO



# Aamun ohjelma

- > Millaisia aineistoja Tietoarkistoon arkistoidaan
- > Arkistointiprosessi
- > Kvantitatiivisen aineiston anonymisointi ja ryhmiin jako
- > Ryhmätyö kvantitatiivisen aineiston anonymisoinnista n. 10 min
- > Ryhmätyön purku
- > Kvalitatiivisen aineiston anonymisointi ja ryhmiin jako
- > Ryhmätyö kvalitatiivisen aineiston anonymisoinnista n. 10 min
- > Ryhmätyön purku
- > Loppukeskustelu

# Aineistot Tietoarkistossa

---

*18.3.2022*

*Emilia Hakkola, tietoasiantuntija*



TIETOARKISTO  
FINLANDS SAMHÄLLSVETENSKAPLIGA DATAARKIV  
FINNISH SOCIAL SCIENCE DATA ARCHIVE





# Yhteiskuntatieteellinen tietoaarkisto (FSD)

- > Tutkimuksen ja opetuksen valtakunnallinen palveluinfrastruktuuri
- > Suomalaisen yhteiskunnan, ihmisten ja kulttuuristen ilmiöiden tutkimiseen kerättyjä sähköisiä aineistoja
- > Pääasialliset asiakasryhmät: opiskelijat, tutkijat ja korkeakouluopettajat
- > Palvelut maksuttomia
- > CTS-sertifioitu (CoreTrustSeal) sähköinen data-arkisto: pitkäaikaissaatavuus = säilytys + hoiva + jakelu
- > [CESSDA](#)n kansallinen palveluntuottaja





# Tietoarkiston perustehtäviä

SÄILYTTÄMINEN

HANKKIMINEN

JAKAMINEN

VAIKUTTAMINEN



TIETOPALVELU

LISÄPALVELUT TIEDOTTAMINEN



## Tietoarkiston palveluita

**Aineistoportaali Aila** – etsi, selaa, lataa ja tallenna tutkimusaineistoja

**Aineistonhallinnan käsikirja**

**Tietopalvelu** – Palvelemme mielellämme myös henkilökohtaisesti: ota yhteyttä!

**Tutkimusmenetelmien verkkokäsikirja**

**Penna** – Kirjoitusaineistojen keruutyökalu

**Pohtiva** – Poliittisten ohjelmien tietovaranto

**Tietoarkistoblogi**





# Tutkimusaineistot Tietoarkistossa



Määrällisiä/kvanti ja laadullisia/kvali



Datatiedosto(t) + metadata = aineisto



Arkistoon tallennettua aineistoa muokataan:  
tavoitteena pitkäaikaissäilytys & jatkokäytettävyys



Ei paperiaineistoja

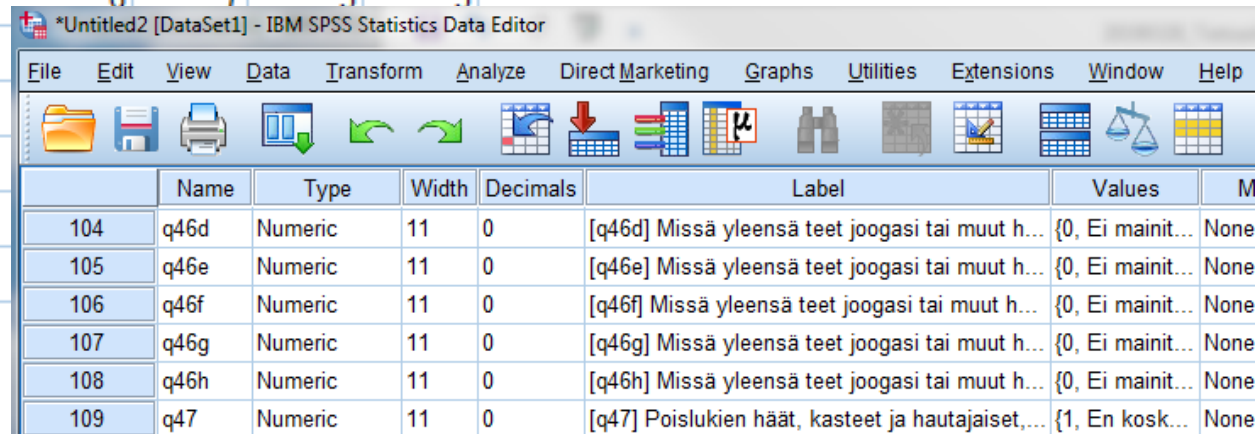


Ei julkaisuja – ts. ei kirjasto

# Kvantitatiiviset eli määrälliset aineistot

- > Kvantidata = tallennettu havaintomatriiseiksi
- > Posti-, puhelin-, internet- tai käyntikyselyjä
- > Havaintoyksikkö voi olla henkilö, organisaatio, kotitalous

q1	q2	q3	q4	q5	q6	q7	q8	q9
2	5	4	5	4	6	2	3	5
3	7	7	4	3	6	4	6	2
2	7	4	5	2	3	3	5	5
5	1	3	4	3	6	7	5	3
2	4	4	2	3				
2	7	2	3	3				
3	7	6	3	3				
1	7	2	3	2				
5	2	6	7	2				



	Name	Type	Width	Decimals	Label	Values	M
104	q46d	Numeric	11	0	[q46d] Missä yleensä teet joogasi tai muut h...	{0, Ei mainit...	None
105	q46e	Numeric	11	0	[q46e] Missä yleensä teet joogasi tai muut h...	{0, Ei mainit...	None
106	q46f	Numeric	11	0	[q46f] Missä yleensä teet joogasi tai muut h...	{0, Ei mainit...	None
107	q46g	Numeric	11	0	[q46g] Missä yleensä teet joogasi tai muut h...	{0, Ei mainit...	None
108	q46h	Numeric	11	0	[q46h] Missä yleensä teet joogasi tai muut h...	{0, Ei mainit...	None
109	q47	Numeric	11	0	[q47] Poislukien häät, kasteet ja hautajaiset,...	{1, En kosk...	None





# Kvalitatiiviset eli laadulliset aineistot

- > Kvalidata = Puhetta, kuvia tai tekstiä sähköisenä tallenteena
- > Esim.
  - Haastattelut
  - Kirjoitukset
  - Havainnointipäiväkirjat
  - Kuvat
  - Piirustukset

18:45 Ohjaaja: Tervehdys! Mihin kaipaat opastusta?

18:47 Chat-kumppani: Hei ! Kaipaen neuvontaa päättämään mikä olisi paras seuraava

18:47 Ohjaaja: Selvä homma. Mikäs on nykyinen tilanteesi?

18:49 Chat-kumppani: Nykyinen tilanteeni on että en tällä hetkellä ole palkka töissä monet alat ja olen yrittäjä henkinen. Kuitenkin olen tällä hetkellä umpikujassa koska valmiiksi.

18:50 Ohjaaja: Okei. Onko sulla pohjakoulutusta? Minkä alan yrittäjyys kiinnostaisi

18:52 Chat-kumppani: Olen käynyt peruskoulun ja vuoden kurssi muotoisesti lukion liikenteen alaa ja yhteiskuntatieteiden, liiketalouden ja hallinnon alaa. Kaikki luotte mielenkiintoisia, mutta työt veivät minua aina mukanaan enkä malttanut aina vuoc jatkaa koulutusta.

18:53 Ohjaaja: oletko miettinyt, että kävisit jonkun koulutuksen loppuun asti? jotta ammatti/tutkinto olemassa.

18:53 Ohjaaja: esimerkiksi oppisopimuksella saisit tehdä työtä samalla

18:54 Chat-kumppani: Yrittäjyys kiinnostaa [henkilöstöalan], [myyntityön] sekä aut kuten pitäen kursseja [liittyen hyvinvointiin]

18:55 Chat-kumppani: Minusta on huonoa sanoa "käy jokin koulutus" vaikka koulu menekkään hukkaan.. on myös tärkeää että siintä tulee tulevaa ammattia ajatellen hyötyä



## Tietoarkiston aineistot

- Laaja kirjo aineistoja, jotka osaltaan tekevät näkyväksi sen, millainen Suomi on ja on ollut.
- Voi tutkia väestön elämää lapsuudesta vanhuusikään.
- Kertovat suomalaisten asenteista, arvoista, koulutuksesta, työelämästä, vapaa-ajasta, sosiaalisista suhteista, osallisuudesta ja osattomuudesta, hyvin- ja pahoinvoinnista.
- Mahdollistavat vertailun.



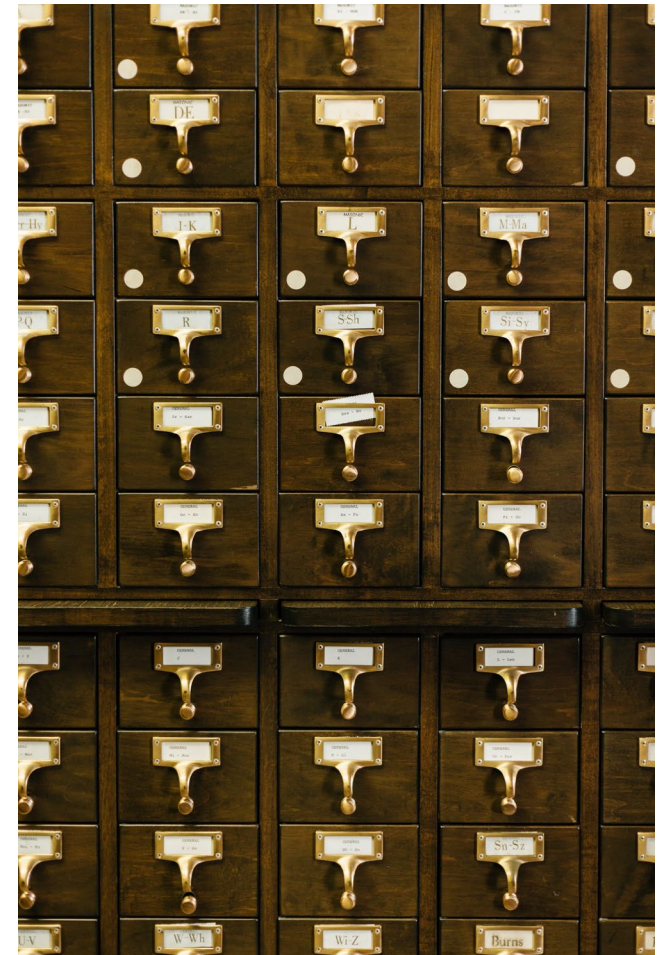


## Mihin ja miksi jo arkistoituja aineistoja?

- Menetelmäopiskeluun, seminaaritöihin, opinnäytteisiin, opetukseen, tutkimukseen
- Tutkimuksen taustatietona / kuvailevina tietoina
- Vertailevaan tutkimukseen tai pitkittäistutkimukseen
- Uuden datankeruun suunnittelun apuna
  
- Vielä paljon tutkittavaa jäljellä
- Uudet teoriat ja menetelmät tuovat uusia mahdollisuuksia
- Tulosten kontrollointi ja toistettavuus
- Useat eettiset kysymykset on jo ratkaistu
- Tutkittavia ei vaivata toistuvasti
- Nopeuttaa valmistumista
- Aikaa ja rahaa säästyy!



# Arkistointiprosessi



# Anonymisointi arkistointiprosessissa

*Arja Kuula-Luumi*

---

*Tietoarkisto*

*Avoim TKIO –verkosto 18.3.2022*



TIETOARKISTO



# Ehdota aineistoa arkistoitavaksi

- > Täytä lomake, joka löytyy arkistointisivulta
- > <https://www.fsd.tuni.fi/fi/palvelut/arkistointi/ohjeita-aineiston-arkistoijalle/>

## Ehdota aineistoa arkistoitavaksi

Voit tarjota aineistoasi arkistoitavaksi. Täytä oheinen lomake ja aineistopalvelumme on sinuun pian yhteydessä. Autamme mielellämme kaikissa aineiston arkistointiin liittyvissä kysymyksissä.

Ehdota aineistoasi arkistoitavaksi 





# Anonymisointi on tutkijan vastuulla

## Poista aineistosta henkilötiedot

Arkistoitavasta aineistosta tulee poistaa tunnistetietoja Tietoarkiston ohjeistuksen mukaan. Poikkeus tähän ovat esimerkiksi lehtiaineistot ja aineistot, joihin perustuviin tutkimusjulkaisuihin sisältyvät tutkittavien henkilötiedot (esimerkiksi asiantuntija- ja taiteilijahaastattelut sekä teossuojan alaiset aineistot).

Yksityiskohtaiset ohjeet tutkimusaineiston anonymisointiin löydät Aineistonhallinnan käsikirjasta [Tunnisteellisuus ja anonymisointi -luvusta](#).

Tietoarkisto auttaa tarvittaessa arkistoitavan aineiston anonymisointisuunnitelman teossa. Tietoarkisto tarkistaa aina alkuperäisen tutkijan jo tekemän anonymisoinnin ja poistaa tarvittaessa lisää tunnistetietoja prosessoinnin aikana.

Anonymisoinnin tulee olla peruuttamaton. Siksi tutkijan tulee hävittää henkilötietoja sisältävät omat aineistotiedostonsa arkistoinnin valmistuttua.





# Anonymisoinnin tasoon eivät vaikuta seuraavat asiat

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





# Anonyymi aineisto

- > Anonymistä 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.

# Kvantitatiivisen aineiston anonymisointi

---

*Emilia Hakkola*

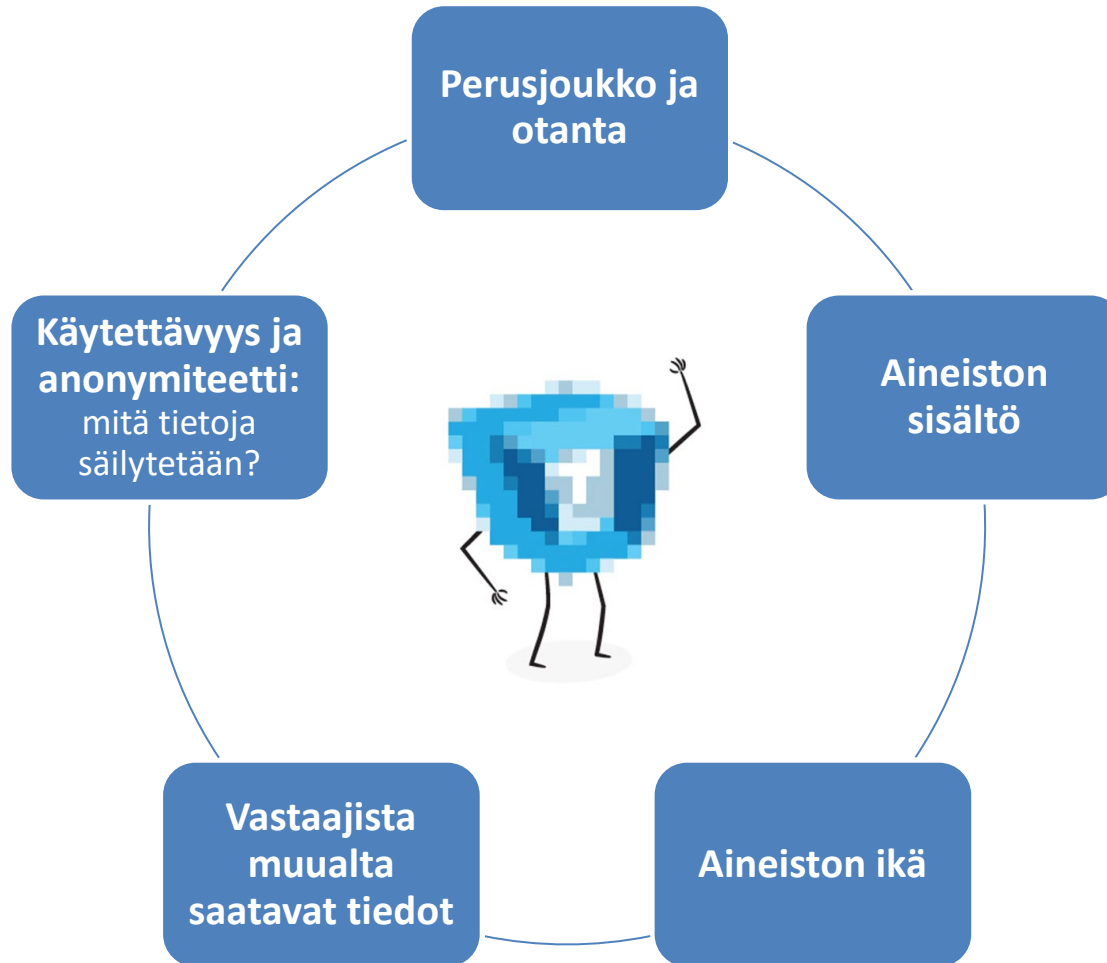
*18.3.2022 Avoin TKIO-verkosto*



TIETOARKISTO



# Mitä asioita tulee ottaa huomioon suunniteltaessa anonymisointia?





## Käytännön anonymisointivinkkejä kvantiaineistoille 1/2

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



## Käytännön anonymisointivinkkejä kvantiaineistoille 2/2

- Ole huolellinen harvinaisten arvojen poistamisessa/muokkaamisessa. Esim. yhtä harvinaista luokkaa ei voi laittaa yksistään SYSMISIin. Eli jos sysmissiin laittaa arvoja, tulee sinne luokitella muitakin harvinaisia erisuuruisia arvoja tai sysmis-arvoissa tulee jo olla alun perin havaintoja.
- Kirjaa tehdyt anonymisointitoimenpiteet esim. syntaksiin, jotta myöhemmin tiedetään, miten alkuperäistä aineistoa on muokattu.
- Käytä avovastausten anonymisointitoimenpiteisiin jotain merkintää, esimerkiksi [hakasulkuja], jotta tiedetään mitä on muutettu ja mitä ei. On silti perusteltua tehdä anonymisointia ilman merkintöjä tilanteissa, missä merkinnästä voi päätellä poistetun asian.



# 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], [yläaste], [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!**



**Aineisto ryhmätyöhön:**

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



# Anonymisointiehdotuksia: kvantitatiivinen aineistoesimerkki 1/2

## Vaihtoehto 1: säilytetään kuntamuuttuja

### LUOKITELLAAN:

- **Syntymävuosi:** luokitellaan 3 vuoden välein
- **Talouden koko:** TOP-koodataan 4 tai yli
- **Sukupuoli** ja **kunta** jätetään sellaisenaan aineistoon

### POISTETAAN:

- Synnyinmaa, ks. huomio alta.
- Harrastusavomuuttuja

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

## Vaihtoehto 2: säilytetään kuntamuuttuja

### LUOKITELLAAN:

- **Kunta:** pienimmät kunnat Hailuoto ja Uurainen yhdistetään
- **Syntymävuosi:** luokitellaan 3 vuoden välein
- **Talouden koko:** TOP-koodataan 4 tai yli
- **Sukupuoli** jätetään sellaisenaan aineistoon
- **Harrasteavomuuttuja:** suurimmat luokat jätetään sellaisenaan, pienimmät yhdistetään

### POISTETAAN:

- Synnyinmaa, ks. huomio alta.
- HUOM. Voidaan tehdä myös niin, että **synnyinmaata ei poisteta vaan** luokitellaan 1 Suomi, 2 muu, **JOS** pienimmissä kunnissa Hailuodossa ja Uuraisilla ei ole tässä aineistossa muualla kuin Suomessa syntyneitä.





# Anonymisointiehdotuksia: kvantitatiivinen aineistoesimerkki 2/2

## Vaihtoehto 3: Ei kuntamuuttujaa!

Kaikki muut tiedot voidaan jättää sellaisenaan paitsi:

1. Synnyinmaa luokitellaan niin, että sinne jää yleisimmät tai Suomi ja muu
2. Avoimesta harrastusmuuttujasta poistetaan lajikerho-maininnat, kilpaurheilun laji poistetaan, s-postiosoitteet, blogimaininnat sekä koirien ja ihmisten nimet

# Kvalitatiivisen aineiston anonymisointi

---

*Arja Kuula-Luumi*

*Tietoarkisto*

*Avoim TKIO –verkosto 18.3.2022*



TIETOARKISTO





# Poista/karkeista suorat tunnisteet ja vahvat epäsuorat tunnisteet

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



# Anonymisoinnin keinoja (Osa 1)

## > Taustatietojen luokittelu

- Arja Kuula-Luumi, johtava asiantuntija Tietoarkistossa, 58 v, asuu Tampereen Lentävänniemessä →
- [nainen] tai [Pirjo], asiantuntija yliopistossa, ikäryhmä: 55-59 vuotias, asuinpaikka: suuren kaupungin lähiö, Etelä-Suomi

## > Henkilönimien vaihtaminen keksityiksi nimiksi

- Arja Kuula-Luumi → [Pirjo]



# Anonymisoinnin keinoja (osa 2)

## > Erisnimien kategorisointi

- Arja → [nainen, naapuri], [nainen, työtoveri], [vaimo], [sisko], [täti]
- SeAMK → [ammattikorkeakoulu], [korkeakoulu]
- Joupin koulu → [alakoulu]
- Kalevan Navetta → [kulttuurikeskus]
- Talriikki → [ravintola]

## > Aineisto-osuuksien poistaminen

## > Aluetietojen luokittelu

- Seinäjoki → [kaupunkimainen kunta, Länsi-Suomi]
- Alavus → [taajaan asuttu kunta, Etelä-Pohjanmaa]



TIETOARKISTO  
FINNISH SOCIAL SCIENCE  
DATA ARCHIVE

## Aineistoesimerkki ryhmätyöhön

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



# Yksi malli anonymisoinnista

- > T: Kerro vähän alkuun itsestäsi ja liitostasi ennen eroa.
  
- > H5: No niinku jo puhelimessa sanoin, olen ollut töissä sairaanhoitajaksi valmistuttuani [terveyskeskuksessa].[Sannin] syntyessä vielä opiskelin [-] ammattikorkeakoulussa. [Saku] valmistui mua myöhemmin [-] yliopistosta ja aluksi asuttiin [tietyn kadun] perheasunnossa, mikä oli tosi hyvä, kun mun äiti jäi just eläkkeelle. Tuskin olisi niin onnistunut, jos ei mummo olisi aina ollut apuna, siis asui siinä [lähellä]. Siinä vaiheessa ei riitoja kyllä ollut, yhtä hyörinää. Kai [Saku]kin ajatteli, että yhdessä elämä edetään.
  
- > T: Missä vaiheessa ero sitten tuli?
  
- > H5: Päätös tehtiin yhdessä vuosi sitten ja muistan sen päivän tosi hyvin.[kesällä 2018] ja jotenkin se iski lujasti, vaikka itsekään vaihtoehtoa nähny. Kirjotin faceen illalla, että tästä päivästä eteenpäin ei enää [Saku]n kanssa ja semmosta kauhistelua, myös kannustusta tuli, mutta en sitten kyllä siellä jatkanut sitä vatvomista. Yhdessä oltiin oltu kahdeksan vuotta ja [Sanni] oli viisi. [Saku] viimeisteli gradua [gradun aihe poistettu] ja sitten ei kai itse halunnut olla itsekään yksinäinen edes opiskeluissa, kun tuli tämä kolmiodraama kuvaan, sulle puhelimessa selitinkin.