

Tutkimusaineiston anonymisointi ja jatkokäyttö tietosuoja-asetuksen valossa

Annika Sallinen

Kierrätä ja käytä –dataseminaari 20.3.2018



TIETOARKISTO



EU:n tietosuoja-asetus ja henkilötiedon määritelmä

- > Sovelletaan 25.5.2018 lähtien henkilötietoa sisältäviin tutkimusaineistoihin
- > Henkilötiedon määritelmä:
 - > Kaikki tunnistettuun tai tunnistettavissa olevaan luonnolliseen henkilöön liittyvät tiedot
 - > Tunnistaminen voidaan tehdä suoraan tai epäsuorasti
 - > Esimerkkejä tunnisteista: nimi, henkilötunnus, sijaintitieto, verkkotunnistetieto (esim. IP-osoite), fyysiset, fysiologiset, geneettiset, psyykkiset, taloudelliset, kulttuurilliset tai sosiaaliset ominaisuudet

Huomaa, että myös ei-ihmisiä (*apersonal*) käsittevä tieto voi olla tunnisteellista (Elliot et al. 2016)



Tietosuoja-asetuksen vaatimuksia henkilötiedon käsittelylle ja jatkokäytölle

- > Henkilötietojen käsittelylle oltava **laillinen käsittelyperuste**, tutkimuksessa se on usein suostumus
- > **Suostumuksen** tulee olla *vapaaehtoinen, tietoinen ja yksilöity*
- > **Käyttötarkoitussidonnaisuus**: aineistoa voi käyttää vain ennaltamäärättyyn tarkoitukseen
- > **Tietojen minimointi**
- > **Erityisten tietoryhmien (ent. arkaluonteisiin) tietojen käsittely**: suostumus ja tietosuojan vaikutusten arviointi
- > Rekisterinpitäjällä on **osoitusvelvollisuus** (informointi, suostumukset, suojatoimet)



Anonyymi vai pseudonyymi?

Anonyymi tieto: Yksittäinen havainto (henkilö) ei ole kohtuullisin keinoin tunnistettavissa annettujen tietojen perusteella tai tietoja muihin tietoihin yhdistämällä.

Anonymisointi ja pseudonymisointi ovat tietosuojasetuksen suojatoimia, jotka mahdollistavat aineiston jatkokäytön.

Pseudonyymi tieto: Yksittäinen havainto ei ole tunnistettavissa pseudonymisoitujen tietojen perusteella ilman ulkopuolisia lisätietoja, jotka prosessin jälkeen säilytetään erillään aineistosta ja suojataan organisatorisesti ja teknisesti. Aineisto on pseudonyymi niin pitkään kuin erillään oleva tieto tuhoetaan. Esim. kohorttitutkimukset



Entä minun aineistoni?

Mistä tiedän sisältääkö aineistoni henkilötietoja, joiden perusteella henkilöt voivat olla tunnisteellisia?

- > **Ei henkilötietoja:** Silloin ei sovelleta GDPR:ää
- > **Henkilötietoa sisältävä data vain omaan käyttöön:** silti tulee noudattaa tietojen minimoinnin periaattetta sekä suojata aineisto teknisesti ja organisatorisesti.
- > Tutkittavien yhteystiedot



Anonymisoinnin suunnittelu

- > Anonymisointi on aina aineistokohtainen
- > Täysin anonyymiä tietoa ei ole olemassa
- > Tieto ihmisistä lisääntyy koko ajan, tee tunnisteellisuuden arviointia säännöllisesti (ns. jäännösriskin arviointi)
- > Tee kirjallinen anonymisointisuunnitelma!





Anonymisoinnin apulista:

1. Tutkittavien informointi ja suostumus
2. Aineiston käyttöoikeus ja saatavuus
3. Aineiston piirteet: kohdejoukko, otantamenetelmä, aineiston koko, aikaulottuvuus, vastausprosentti
4. Aineiston suorat ja epäsuorat tunnisteet
5. Sisällön sensitiivisyys
6. Muualta saatavat, aineistoon yhdistettävät tiedot
7. Anonymisointipäätökset
8. Anonymisoinnin perustelut (käytettävyys vs. anonymisointi)
9. Paljastumisriskin arviointi





Paljon tietoa aineiston- hallinnan käsikirjassa!

1 Miksi aineistonhallintaa ja jatkokäyttöä?

- Osa hyvää tieteellistä käytäntöä
- Rahoittajat edellyttävät aineistonhallintaa
- Jatkokäytön ja avoimen saatavuuden hyödyt
- Tutkimusaineiston elinkaari

2 Aineistonhallinnan suunnittelu

- Aineistonhallintasuunnitelma
- 1. Aineisto
- 2. Oikeudet
- 3. Tietosuojat ja tietoturva
- 4. Tiedostoformaatit ja ohjelmistot
- 5. Aineiston käsittelyn ja sisällön kuvaus
- 6. Elinkaari

3 Sopimukset ja oikeudet

- Tekijänoikeus
- Tutkijoiden välisistä sopimuksista
- Oikeuksien siirtymisestä sopiminen
- Tekijyys tieteellisissä julkaisuissa ja tutkimusaineistoissa

4 Tutkittavien informointi

- OSA I: Informoinnin merkitys
- OSA II: Informoinnin sisältö, muoto ja tutkittavan suostumus
- OSA III: Informoinnin osa-alueet
- Osa IV: Esimerkkejä informoinneista

5 Tiedostoformaatit ja ohjelmistot

- Tiedostojen konvertointi
- Tutkimusaineiston digitointi
- Havaintomatriisit ja tilasto-ohjelmistot
- Tekstiaineistot
- Kuva-aineistot
- Ääni- ja äänikuvatallenteet

6 Kvantitatiivisen datan käsittely

- Havaintomatriisin tallennus
- Muuttujien nimet ja selitteet
- Muuttujan arvot ja niiden selitteet
- Muuttujien muunnokset
- Puuttuva tieto
- Painomuuttujat
- Syntaksin käyttö

7 Kvalitatiivisen datan käsittely

- Litterointi
- Aineiston sisältämien datatiedostojen organisointi
- Datatiedostojen nimeäminen
- Taustatietojen kirjaamistavat
- Lehtiaineistot

8 Tunnisteellisuus ja anonymisointi

- Mitä on henkilötieto?
- Milloin tieto on tunnistettava?
- Milloin tieto on anonymiä?
- Tunnisteellisten aineistojen käsittely
- Anonymisoinnin lähtökohdat
- Kvantitatiivisen aineiston anonymisointi
- Kvalitatiivisen aineiston anonymisointi

9 Aineiston kuvailu ja metadata

- Metadata tallennus
- Tutkimuksen toteutuksen kuvailu
- Aineistonkeruunstrumentti
- Tiedostojen kuvailu
- Muuttujien kuvailu
- Saatavuustiedot
- Kontekstitiedot ja paradata
- Metadataformaatit

10 Fyysinen säilytys

- Tiedostojen nimeäminen ja hallinta
- Varmuuskopiointi
- Migraatio ja virkistäminen
- Tallennusmediat
- Tietoturva
- Aineiston hävittäminen
- Paperiaineiston säilytys

11 Malliesimerkit ja sanastot

Sisällys





Lähteet:

- > Elliot, Mark & Mackey, Elaine & Kieron O'Hara & Tudor, Caroline (2016). [The Anonymisation Decision-Making Framework](#). Manchester: UKAN.
- > GDPR in EUR-lex: <http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679>
- > Tarhonen, Laura (2016). [Pseudonymisation of Personal Data According to the General Data Protection Regulation](#). Viestintäoikeuden vuosikirja 2016, 10–32.
- > Pictures: www.unsplash.com



Esitä
rohkeasti
kysymys!

Yhteystiedot:
www.tietoarkisto.fi
asiakaspalvelu.fsd@uta.fi



You Tube

tietoarkisto



TIETOARKISTO