Täydentävä ohje arkaluonteisia ja luottamuksellisia tietoja sisältävän datan hallinnan suunnitteluun 2019

Viittausohje: ATT aineistonhallinnan ohje sensitiivisille aineistoille - työryhmä (2019): Ohje arkaluonteisia ja luottamuksellisia tietoja sisältävän datan hallinnan suunnitteluun, Tuuliprojekti.
Zenodo: 10.5281/zenodo.3247282

Johdanto	
	Tämä ohje koskee aineistotyyppejä, joissa on mukana arkaluonteista ja luottamuksellista tutkimusaineistoa. Tutkimusorganisaatiot opastavat tutkijoita tietosuoja- ja tietoturvaperiaatteiden noudattamisessa. Tämän lisäksi organisaation DMP-yleisohjeesta löytyvät organisaatiokohtaiset muut keskeiset tukipalveluiden yhteystiedot. Organisaatiot ovat myös voineet yhdistää tämän ja yleisohjeen. Tutustu organisaatiosi aineistonhallintaohjeisiin.
1. Aineiston yleiskuvaus	
1.1 Millaiseen aineistoon tutkimuksesi perustuu? Millaista aineistoa kerätään, tuotetaan tai käytetään uudelleen? Missä tiedostomuodoissa aineisto on? Arvioi myös karkealla tasolla, kuinka paljon levytilaa aineistosi lopulta tarvitsee?	Erittele kaikki aineistotyypit, jotka sisältävät henkilötietoa, arkaluonteista tai luottamuksellista tietoa. Tutkimusaineiston arkaluonteiset osat on erityisen tarpeellista tunnistaa, koska aineistonhallinnan suunnittelussa keskitytään näihin liittyvien riskien tunnistamiseen ja hallintaan. Henkilötietojen osalta kerro, mikä taho toimii *rekisterinpitäjänä.* Arkaluonteinen ja luottamuksellinen tieto on sellaista, joka voi paljastuessaan aiheuttaa vahinkoa: • Arkaluonteinen henkilötieto; arkaluonteisista henkilötiedoista ei voi tehdä kattavaa listausta. **Tutkimuksen tekijöiden vastuulla on tunnistaa tiedot, joiden paljastumisesta saattaisi olla haittaa tutkittaville.** ○ Arkaluonteiset tiedot voivat liittyä terveyteen tai sairastumisriskeihin, seksuaaliseen suuntautumiseen, etniseen alkuperään, ammattiliittoon kuulumiseen, uskonnollisiin vakaumuksiin. • Sensitiivinen lajitieto, kuten uhanalaiset eläimet ja kasvit, luonnonsuojelullinen tieto tai bioturvallisuusuuteen liittyvä tieto. • Muu luottamuksellinen tieto, kuten patentit, maanpuolustukseen liittyvä tieto, organisatorinen tieto tai liikesalaisuudet.

	Vinkkejä Henkilötietoja ovat kaikki ne tiedot, joista henkilö on yksilöitävissä joko *suoraan* tai *epäsuorasti*. • *Suorat tunnisteet*: nimi, puhelinnumero, henkilötunnus, kuva, ääni, sormenjälki, hammaskartta, MRI-kuva • *Epäsuorat tunnisteet*: sukupuoli, ikä, koulutus, ammattiasema, kansallisuus, sijaintitunnisteet, työhistoria, järjestelmän lokitiedot, sivilisääty, asuinpaikka, auton rekisterinumero Linkkejä: Mitä on henkilötieto (FSD), Henkilötietojen käsittely (Tietosuojavaltuutetun toimisto)
1.2 Miten aineiston yhtenäisyys ja laatu varmistetaan?	Pohdi, miten mahdollinen minimointi, pseudonymisointi tai anonymisointi vaikuttavat aineiston laatuun. https://www.fsd.uta.fi/aineistonhallinta/fi/tunnisteellisuus-ja-anonymisointi.html#milloin-tieto-on-anonyymia-enta-pseudonyymia https://tietosuoja.fi/pseudonymisointi-anonymisointi
2. Eettisten periaatteiden ja lainsäädännön noudattaminen	
2.1 Mitä eettisiä seikkoja aineistosi hallintaan liittyy (esim. arkaluonteisten tietojen käsittely, tutkittavien identiteetin suojaaminen ja tietojen jakamista koskevan suostumuksen hankkiminen)?	Tarkista: *rekisterinpitäjyys* Erityisten henkilötietoryhmien käsittely: https://tietosuoja.fi/erityisten-henkilotietoryhmien-kasittely • Milloin erityisiä henkilötietoryhmiä saa käsitellä? https://tietosuoja.fi/erityisten-henkilotietoryhmien-kasittely Tietosuojavaltuutetun toimisto: usein kysyttyä: https://tietosuoja.fi/usein-kysyttya https://www.tenk.fi/fi/ihmistieteiden-eettinen-ennakkoarviointiohje-uudistuu Kansallinen neuvottelukunta TUKIJA: Valtakunnallinen lääketieteellinen tutkimuseettinen toimikunta Katso oman organisaatiosi ohjeet.
2.2 Miten aineiston omistajuuteen, tekijänoikeuksiin ja immateriaalioikeuksiin liittyviä asioita hallintaan? Estävätkö tekijänoikeudet, käyttöoikeudet tai muut	Aineiston omistajuutta sekä muita immateriaalioikeuksia koskevat sopimukset tehdään ennen varsinaista konkreettisten tutkimustoimien aloittamista. Sopimusmallit sekä konsultaatio on järjestetty omassa tutkimusorganisaatiossasi.

rajoitukset aineiston käyttämisen tai jakamisen?	
3. Dokumentointi ja metatiedot	
3.1 Miten dokumentoit aineistosi, jotta se on löydettävissä, saavutettavissa, yhteentoimivaa ja uudelleen käytettävissä sekä itseäsi että muita varten? Mitä metatietostandardeja, README-tiedostoja ja muuta dokumentaatiota käytät, jotta muut voivat ymmärtää ja käyttää aineistoasi?	Muista aineiston kuvailun yhteydessä, että myös tiedostonimet, tiedostokansioiden nimet sekä muuttujat ja metadata saattavat sisältää henkilötietoja tai arkaluonteista tietoa. Vaikka tutkimusdatasi sisältäisi henkilötietoja, voi metadatan julkaista, jos metadata ei sisällä tunnisteellista tietoa, jonka avulla tutkittavan voi identifioida. • Making a research project understandable - Guide for data documentation DOI 10.5281/zenodo.1683181 • Aineistonhallinnan käsikirja
4. Tallentaminen ja varmuuskopiointi tutkimushankkeen aikana	
4.1 Minne aineistosi tallennetaan ja miten se varmuuskopioidaan?	Tietosuoja-asetuksen mukaan henkilötietojen käsittelyyn liittyviä riskejä on arvioitava ennen kuin henkilötietoja ryhdytään käsittelemään. Tutustu organisaatiosi tietosuoja- ja riskienhallinnan ohjeisiin ja mieti: • Mitä rekisteröidyn vapauksia ja oikeuksia käsittely voi vaarantaa? • Mitä vahinkoja rekisteröidylle voi aiheutua suunnitellusta henkilötietojen käsittelystä? • Mitä vahinkoja rekisteröidylle voi aiheutua, jos aineisto päätyy vääriin käsiin, tuhoutuu tai pilaantuu? • Millaisilta riskeiltä aineistojasi on suojattava? • Millä keinoin tunnistettuja riskejä hallitaan? • Mikä on jäännösriskien todennäköisyyksille ja vaikutuksille hyväksyttävä taso? Arvion jälkeen varmista organisaatiosi tietosuojavastaavalta edellyttääkö aineistosi tietosuoja-asetuksen mukaista vaikutustenarviointia. Selvitä myös, onko jollakin ulkoisilla tahoilla, kuten tutkimuksen rahoittajalla tai aineiston omistajalla omia vaatimuksiaan aineistoon liittyen. Riskiarvion perusteella määritellään suojatoimenpiteet aineiston koko elinkaaren ajaksi (Ks. myös kohta **4.2**).

	Mieti:
	
	• Mitä tallennuspalveluja ja -laitteita tutkimuksen aikana käytetään?
	• Ketkä vastaavat käytettyjen tallennuspalvelujen ylläpidosta?
	• Miten varmuuskopiointi tehdään käyttämissäsi tallennuspalveluissa?
	  ○ Kuka vastaa varmuuskopioinnin toteuttamisesta?
	  ○ Minne varmuuskopiot tallennetaan?
	  ○ Kuinka usein varmuuskopioita otetaan?
	  ○ Kuinka kauan varmuuskopioita säilytetään?
	• Pitävätkö tallennuspalvelut kirjaa (lokia) aineiston käytöstä?
	• Onko aineisto etäkäytettävissä?
	  ○ Jos kyllä, miten etäkäyttö suojataan?
	• Tarvitaanko tietojen salausta?
	  ○ Jos kyllä, mieti:
	  ▪ Mikä osa aineistosta salataan ja mitä ei?
	  ▪ Mitä salausvälineitä käytetään?
	  ▪ Kuka hallinnoi salausavaimia ja salasanoja?
	• Miten aineiston käsittelyyn käytetyt tilat on suojattu?
	  ○ Saako työtilojen ovet lukkoon?
	  ○ Tunnetaanko kaikki kulkuoikeuksien haltijat?
	  ○ Onko kiinteistössä nauhoittava kameravalvonta?
	  ○ Onko käytettävissä murtosuojattuja säilytyskalusteita tai -tiloja fyysisille aineistoille ja tallennuslaitteille?
	  ○ Onko työpisteet näkösuojattu?
	• Miten poistettavat aineistot ja kopiot hävitetään turvallisesti kun niiden käyttötarve päättyy?
	
	Tutustu oman organisaatiosi ohjeeseen tallennuspalveluista ja työvälineistä, joilla taataan aineistojen tietoturvallinen käsittely. Selvitä myös organisaatiosi tietoturvayksikön ja IT-yksikön palveluosoite.
	
	Lisätietoja:
	
	Riskien arviointi (Tietosuojavaltuutetun toimisto)
	
	Vaikutustenarviointi (Tietosuojavaltuutetun toimisto)

4.2 Kuka valvoo pääsyä aineistoon ja miten suojattua pääsyä aineistoon valvotaan?	Pääsy henkilötietoihin on rajattava koskemaan vain niitä henkilöitä, jolle se on tutkimuksen suorittamisen kannalta tarpeellista. Ota huomioon, että kyseinen joukko sisältää myös käyttämiesi palvelujen ja välineiden ylläpitäjät ja muut, mahdollisesti ulkopuoliset palvelutoimittajat. Mieti:

	• Miten, kenen ja millä tavalla aineistoa tarvitsee käyttää?
	  o Kenelle käyttöoikeus voidaan myöntää?
	  o Millainen käyttö on sallittua?
	  o Mikä osa aineistoa edellyttää minkäkinlaisia käyttöoikeuksia?
	  o Tarvitseeko aineistoa jakaa yhteistyökumppanien tai palvelutoimittajien kanssa?
	  o Kuka ja millä perustein voi siirtää aineistoa osapuolelta toiselle?
	• Miten käytön- ja kulunvalvonta on toteutettu?
	  o Kuka on pääsyoikeuksista vastaava henkilö?
	  o Ovatko käyttö- ja kulkuoikeudet selvillä, muokattavissa ja poistettavissa?
	  o Tarkistetaanko oikeuksien peruste ja rajaus säännöllisesti?
	• Miten aineiston käyttöä ja käytön asiallisuutta valvotaan?
	  o Kuinka usein aineiston käyttöoikeudet tarkistetaan?
	  o Mihin ja kenelle aineistoa tai sen osia kopioidaan?
	  o Miten aineistokopioita hallitaan?
	  o Miten aineistokopiot poistetaan kun käyttöoikeus päättyy?
	  o Miten varmistutaan aineiston käyttötarkoituksista?
	• Onko aineistoa käsittelevien henkilöiden tietosuoja- ja -turvaosaaminen ja käsittelyä koskeva ohjeistus ajan tasalla?
	
	Tutustu oman organisaatiosi pääsyoikeuksien hallinnan periatteisiin, ohjeisiin ja välineisiin. Selvitä miten organisaatiossasi ilmoitetaan henkilötietoihin liittyvistä väärinkäytöksistä ja vahingoista.

5. Aineiston avaaminen, julkaiseminen ja arkistointi tutkimushankkeen päätyttyä	
5.1 Mikä osa aineistosta voidaan asettaa avoimesti saataville tai julkaista? Missä ja milloin aineisto tai siihen liittyvät metatiedot asetetaan saataville?	Henkilötietoja sisältävät aineistot on mahdollista avata ainoastaan anonymisoituina. Anonymisointi kannattaa muutoinkin, sillä anonymisoitu aineisto ei ole enää henkilötietoa eikä siten tietosuojalainsäädännön piirissä ja sen avaaminen ja jakaminen on mahdollista. Pseudonymisoitu aineisto on edelleen henkilötietoa ja sitä ei tästä syystä voi avata. Henkilötietoja sisältävä data voidaan kuitenkin jakaa siitä kiinnostuneiden kanssa luvanvaraisesti alkuperäisen käsittelyperusteen mukaiseen tarkoitukseen. Henkilötietoja sisältävän aineiston alkuperäinen käsittelyperuste, esimerkiksi lakiin perustuva tai suostumus, voi rajoittaa aineiston jatkokäyttöä. Jos esimerkiksi alkuperäisessä suostumuslomakkeessa ei ole huomioitu aineiston jatkokäyttöä, aineiston avaaminen voi vaatia uuden suostumuksen hankkimista tutkittavilta. Henkilötietoja sisältävän aineiston avaamiseen tai julkaisemisen tapoja ovat:

	1. Datan anonymisointi ja anonymisoidun datan avaaminen data-arkistossa. 2. Datan keskeisten metatietojen julkaiseminen (tutkimustietojärjestelmässä tai muussa julkaisupalvelussa) ja varsinaisen aineiston saattaminen käyttöön luvanvaraisesti aineiston tuottajalta tai luotettavasta datarepositoriosta. *Linkkejä* • Pseudonymisoidut ja anonymisoidut tiedot: https://tietosuoja.fi/pseudonymisointi-anonymisointi • Tunnisteellisuus ja anonymisointi: http://www.fsd.uta.fi/aineistonhallinta/fi/tunnisteellisuus-ja-anonymisointi.html • Metatietojen julkaisupalvelu Etsin: https://etsin.fairdata.fi/
5.2 Mihin pitkällä aikavälillä arvokkaat tiedot arkistoidaan ja kuinka pitkäksi ajaksi?	Opetus- ja kulttuuriministeriö tarjoaa korkeakouluille ja tutkimuslaitoksille palvelun tutkimusaineistojen pitkäaikaissäilytykseen (ms. PAS-palvelu). Kukin organisaatio määrittää prosessinsa pitkällä aikavälillä arvokkaiden tutkimusaineistojen tunnistamiseksi jasiirtämiseksi PAS-palveluun. Riippuen organisaation ohjeistuksesta ja tutkimusluvasta PAS-palveluun on mahdollista tallentaa myös arkaluonteisia henkilötietoja sisältäviä aineistoja. Arkaluontoista henkilötietoja sisältävän aineiston arkistointi vaatii säilytysluvan Kansallisarkistolta. Aineisto on minimoitava ennen säilytystä. Tällaisen aineiston jatkokäyttö on mahdollista tutkimusluvalla. Perinteisesti arkaluonteinen aineisto kehotetaan hävittämään tutkimushankkeen jälkeen, koska sen säilyttäminen on riskialtista ja vaatii erityisjärjestelyjä. Siksi on tärkeää myös suunnitella, miten aineisto hävitetään turvallisesti. Esimerkiksi pelkästään tiedoston poistaminen (deletointi) ja tietokoneen roskakorin tyhjentäminen ei tarkoita, että tiedosto olisi tuhoutunut lopullisesti. Poistettuja tietoja voi palauttaa jopa kiintolevyn uudelleen alustamisen jälkeen. Tiedon lopulliseen hävittämiseen on olemassa erilaisia ohjelmia, jotka perustuvat esimerkiksi tietojen ylikirjoittamiseen tai kiintolevyn magnetointiin. Tallennusväline voidaan myös murskata mekaanisesti lukukelvottomaksi. Anonymisoinnin seurauksena aineisto ei sisällä enää henkilötietoja eikä siten ole tietosuojalainsäädännön piirissä. *Vinkkejä* • Muistathan, että aineiston anonymisointi ja hävittäminen tai arkistointi tehdään viimeistään tutkimusluvan määräajan päättyessä.

	• Aito anonymisointi edellyttää sekä suoran että välillisen tunnistamisen mahdollisuuden poistamista sekä tunnisteavaimen hävittämistä. • Useilla korkeakouluilla ja tutkimuslaitoksilla on oma sisäinen ohjeistuksensa tallennusvälineiden hävittämisestä. *Linkkejä* • Tutkimusaineiston hävittäminen: http://www.fsd.uta.fi/aineistonhallinta/fi/fyysinen-sailytys.html#havittaminen
6. Aineistonhallinnan vastuut ja resurssit	
6.1 Ketkä vastaavat aineistonhallintaan liittyvistä tehtävistä tutkimusprojektin elinkaaren aikana? Arvioi myös aineistonhallintaan tarvittavat resurssit (esim. taloudelliset, ajalliset, työmäärään liittyvät).	Kuka on **vastuussa** arkaluonteisen ja luottamuksellisen aineiston hallinnasta ja sen toteutumisen valvonnasta aineiston koko elinkaaren ajan? • Kuka vastaa *tietosuojasta* (katso kohta 2) ja *tietoturvasta* (katso kohta 4)? Tarvittavia **resursseja** suunniteltaessa ota huomioon: • aineiston minimoinnin, pseudonymisoinnin ja anonymisoinnin kustannukset eli siihen menevä aika ja tarvittavat ohjelmistot. • korkeamman turvatason vaatimukset toiminnalle ja tekniikalle sekä niistä koituvat lisäkustannukset.