

## **DMP Tuuli – Aineistohallintasuunnitelma Kareliassa (27.8.2019)**

Viittaus Karelian datapolitiikkaan: koskee kaikkia karelijalaisia

### OSIO 1: AINEISTON KERÄÄMINEN JA DOKUMENTAATIO

#### Kysymys 1:

#### **Millaista aineistoa hankkeessa /opinnäytetyössä kerätään tai uudelleen käytetään?**

Kuvaile lyhyesti, millaista aineistoa olet keräämässä tai mitä olemassa olevaa aineistoa aiot hyödyntää. Kerro, miten aineisto tullaan keräämään ja minkä tyyppistä aineisto on (esim. kysely, haastattelu, havainnointi, kuva/ääni, koneellinen mittaustulos, mallinnus, fyysinen näyte, ym.).

#### **Hyviä käytänteitä**

- Aineistohallintasuunnitelmassa ei tarvitse kuvata tutkimusmenetelmiä uudelleen, vaan voit viitata tutkimus- tai hankesuunnitelmaan
- Hyödynnä yleisesti hyväksyttyjä aineistonkeruumenetelmiä, jos alallasi niitä on
- Kun käytät uudelleen jo kerättyä aineistoa, vältät kaksinkertaisen työn
- Selvitä, onko tutkimuksessasi tarvetta eettiselle ennakko-arvioinnille
- Muista informoida tutkimuksesi kohdejoukkoa kerättävän aineiston käytöstä (ml. muut käyttötarkoitukset kuin vain kyseinen työ)
- Jos joku tutkimuksen kohdehenkilö kieltää aineiston käytön kesken projektin, siihen asti kerätty aineisto on käytettävissä
- Harkitse aineistosi jakamista ja avaamismahdollisuutta jo tutkimuksen alussa
- Tutustu hyvän tieteellisen käytännön, tietosuojalain ja henkilötietolain peruseriaatteisiin

#### Mallivastaus:

*Tiedot kerätään posti-/sähköisellä kyselyllä. Kysymykset kattavat [tähän tutkimuksen keskeiset aihealueet]. Kutsu kyselyyn lähetetään edustavalle otokselle, joka käsittää 2000 suomen- ja ruotsinkielistä, iältään 39-79 vuotiasta maaseudulla asuvaa suomalaista. Kutsussa on mukana kohdejoukon suostumuslauseke. Vastausprosentiksi arvioidaan 20-40 %. Aineiston hallinnassa hyödynnetään [tähän ohjelmiston nimi].*

#### Kysymys 2:

#### **Mitä tallennusformaatteja käytetään?**

Tallennusformaattien valintaan tulee kiinnittää huomiota, jotta aineiston käytettävyys säilyisi mahdollisimman pitkään. Luettele käyttämäsi tiedostomuodot (esim. .csv, .txt, .mp3).

#### **Hyviä käytänteitä**

- Luettele dataformaatit sekä ohjelmistot, joita niiden käyttämiseen tarvitaan
- Käytettävyuden varmistamiseksi on suositeltavaa tallentaa yksi tiedostokopio sellaisessa muodossa, joka on mahdollisimman yleisesti käytössä ja jota useimmat eri ohjelmistot tukevat, esim. .txt, .jpeg, .mp4
- Käytä avoimen lähdekoodin ohjelmia ja formaatteja, jotta mahdollistat aineiston uudelleen käytön, yhteiskäytön ja jakamisen

#### Mallivastaus:

*Digitaaliset videotiedostot ovat MPEG-4 (.mp4) –muodossa. MPEG-4 on ISO-standardin mukainen ja se on avattavissa useimmilla mediasoitinohjelmilla.*

### Kysymys 3:

#### **Miten aineisto dokumentoidaan ja miten sen eheys ja laatu varmistetaan?**

Aineiston dokumentaatio esittelee käytetyt termit, muuttujien nimet, koodit ja lyhenteet. Mieti millaisia tietoja tarvitaan, jotta aineisto on löydettävissä, käytettävissä ja tulkittavissa myös tulevaisuudessa. Kuvaile millaista dokumentaatiota tarvitaan aineiston liitteeksi.

Aineiston laadunhallinnalla varmistetaan, että aineisto ei katoa tai muutu vahingossa tutkimuksen/hankkeen aikana. Laadunhallinta on oleellinen osa tutkimusta ja siitä tulee huolehtia aineistoa kerätessä, tallennettaessa sekä tarkistettaessa.

#### **Hyviä käytänteitä**

- Mieti miten aineisto organisoidaan projektin aikana kuvaamalla esim. nimeämiskäytännöt, versionhallinta ja kansiorakenne.
- Kerro miten aineiston keruumenetelmä vaikuttaa aineiston laatuun. Muista, että aineiston huolellinen dokumentointi on oleellista laadun hallinnan kannalta.
- Laaduntarkkailuun liittyvät toimenpiteet voivat sisältää esim. standardoituja menetelmiä ja ohjeita havainnoinnin toteuttamiseen, automaattisesti tallentavia lomakkeita selkeillä käyttöohjeilla, ohjeistuksia toistaa mittauksia, havainnoiteja tai näytteiden ottoja järjestelmällisesti tai suorittamalla mittalaitteiden kalibrointia tarpeeksi usein.
- Tutustu tarkemmin tutkimusmenetelmiin esim. [www.vastuullinentiede.fi](http://www.vastuullinentiede.fi) –sivustolla

#### **Mallivastaus:**

*Tässä projektissa kerätyt aineistot x ja y dokumentoidaan nimeämällä ne seuraavasti juoksevilla numeroinnilla: aineisto1001, aineisto1002, aineisto1003...*

*Määrällisen aineiston osalta varmistetaan, että puuttuvat arvot määritellään, muuttujien arvot ovat sallituissa rajoissa ja aineistossa ei ole määrittelemättömiä arvoja.*

\*\*\*

#### **OSIO 2: TALLENTAMINEN JA OMISTAJUUS**

### Kysymys 4:

#### **Miten aineisto tallennetaan ja varmuuskopioidaan?**

Varmista että data on tallennettu kunnolla niin tutkimusprosessin aikana kuin sen jälkeen ja että siitä tehdään tarvittavat varmuuskopiot. Ota huomioon aineiston mahdollinen uudelleenkäyttö.

Mieti kuka on vastuussa aineiston varmuuskopioinnista ja tarvittaessa palauttamisesta. Luo suunnitelma yhdessä mahdollisten yhteistyökumppaneidesi kanssa varmistaaksesi turvallisen tiedonsiirron osapuolten välillä.

Kareliassa kehoitetaan suosimaan työstön aikaisena tallennuspaikkana projektien /henkilöiden omia verkkokansioita (luna), jotka varmuuskopioidaan säännöllisesti organisaation toimesta. Pitkäaikaissäilytyksen paikkoina ovat Karelia-PAS –verkkokansio, mahdolliset erilliset pilvipalvelimet (erityisesti tutkimusympäristöjen suuret data-aineistot), sekä kansalliset ja kansainväliset palvelut kuten IDA, AILA ja EU Open Science Cloud.

## Hyvät käytänteet

- On suositeltavaa käyttää korkeakoulun hallinnoimaa tallennustilaa oman esim. kiintolevyn tai USB-tikun sijaan. Mieti aineiston säilytystä opinto-oikeutesi päättymisen jälkeen.
- Arvioi kuinka paljon tallennustilaa tarvitset
- Onko tallennustilaa tarpeeksi vai tarvitsetko hankebudjettiin varoja lisäpalvelujen hankintaa varten? Muista huomioida asia hankebudjetissa
- Huomioi, että suojatun verkkolevyn käyttö on turvallisempi vaihtoehto kuin kovalevyn käyttö
- Huomioi aineiston tuhoaminen tutkimuksen jälkeen, jos aineistoa ei avata pitkäaikaissäilytykseen

Mallivastaus:

*Aineisto varmuuskopioidaan päivittäin tarkoitukseen varatun salasanan avulla suojatulle turvalliselle palvelimelle, jota ylläpitää organisaatio X.*

### Kysymys 5:

#### **Miten aineiston käyttöoikeuksista huolehditaan?**

Tutkimusprosessin aikana on tärkeää varmistaa, että aineisto säilyy turvassa ja muuttumattomana. Kannattaa määritellä, ketkä kaikki pääsevät aineistoon käsiksi ja millaiset muokkausoikeudet heillä on. Ulkopuolisten ihmisten pääsy aineistoon voi myös olla laitonta. Käyttöoikeudet tulee aina riippua siitä, millaista ja kuinka luottamuksellista aineisto on.

Kareliassa pitkäaikaissäilytetyn aineiston käyttöoikeudet on rajattu erikseen määriteltäville henkilöille.

## Hyvät käytänteet

- Mieti, kuka pääsee käyttämään ja lukemaan aineistoa?
- Miten huolehdit pääsyoikeuksista niin, että data pysyy asianmukaisesti suojattuna?
- Mieti myös toimintamalli siihen, jos ongelmia syntyy (riskienhallinnan suunnitelma)
- Huomioi oikeuksiensiiirto ennen tutkimuksen aloittamista (Karelian mallisopimukset)

Mallivastaus:

*Aineiston analysoinnin aikana siihen on pääsy vain projektiryhmän varsinaisilla jäsenillä. Aineistosta poistetaan kaikki suorat tunnisteet ennen arkistointia.*

### Kysymys 6:

#### **Kuinka hallinnoit eettisiä asioita?**

Kerätessä ihmisiin kohdistuvaa aineistoa, tulee noudattaa tutkimuseettisiä periaatteita ja henkilötietojen käyttöön ja käsittelyyn liittyvää lainsäädäntöä. Mieti miten turvaat yksityisyyden suojan ja miten aineisto anonymisoidaan. Pyydä haastateltavilta tarvittavat luvat ennen aineiston keräämistä. Luvan tulee kattaa koko tutkimusprosessin elinkaari ja siinä tulee ottaa huomioon myös aineiston jakaminen ja jatkokäyttö.

Ennen tutkimuksen aloittamista kannattaa tutustua Tutkimuseettisen neuvottelukunnan (TENK) näkemyksiin sekä [www.vastuullinentiede.fi](http://www.vastuullinentiede.fi) -sivustoon. Huomioi myös lupakäytänteet ja oikeuksien siirto osana sopimuksia ja tiedonkeruuta.

## Hyvät käytänteet

- Selvitä, pitääkö tutkimussuunnitelmasi hyväksyttää eettisessä toimikunnassa. Karelialla on oma eettinen toimikunta, mutta jotkin ihmistieteelliset tutkimukset vaativat myös sairaanhoitopiirin eettisen toimikunnan hyväksynnän.
- Kuvaile, miten ohjeistusta noudatetaan (esim. aineistosta poistetaan henkilötiedot tai muu arkaluonteinen tieto, jotta yksityisyyden suoja säilyy ennen datan jakamista).
- Tutkimusaineiston anonymisointi on osa tutkimuksen eettisyyttä.

Katso lisäohjeita [Tutkimuseettisen neuvottelukunnan](#) sivuilta.

**Mallivastaus:**

*Hankkeessa kerätty data voidaan julkaista ilman rajoituksia, koska data ei sisällä tunnistettavaa tietoa / informanteilta tullaan hankkimaan lupa tutkimusaineiston julkaisemiseen /aineisto anonymisoidaan ennen julkaisemista.*

Kysymys 7:

### **Kuinka hallinnoit tekijänoikeuksia sekä IPR:iä?**

Määrittele, kuka omistaa aineiston ja antaa oikeuden sen jatkoikäkäyttöön. Jos käytät kolmannen osapuolen aineistoa, huolehdi käyttöoikeuksista (copyright) ja lisensoinnista. Karelialla on keskeisistä sovittavista asioista olemassa mallisopimusohjat.

### **Hyvät käytänteet**

- Noudata rahoittajan, organisaation tai osaamisalan käytäntöjä tekijänoikeus- ja IPR-kysymyksissä
- On suositeltavaa, että TKI-hankkeissa tuotetut tutkimusaineistot ovat uudelleen käytettävissä esim. creative commons (CC)-lisenssien avulla. Avoimille aineistoille suositellaan CC-BY lisenssiä.

**Mallivastaus:**

*Tutkimuksessa ei käytetä mitään aineistoa, joka on suojattu tekijänoikeuksilla, mallioikeuksilla tai patenteilla tai muilla vastaavilla menetelmillä. Kaikki hankkeeseen osallistuvat yhteistyökumppanit allekirjoittavat sopimuksen, jonka perusteella aineisto avataan aina kun se on mahdollista. Tekijänoikeus säilyy tutkijoilla, mutta se siirretään myös organisaation haltuun oikeuksiensiirtosopimuksella.*

\*\*\*

## **OSIO 3: AINEISTON AVAAMINEN JA PITKÄAIKAISSÄILYTYS**

Kysymys 8:

### **Miten ja koska aineisto on muiden hyödynnettävissä?**

Kuvaile, onko aineisto avattavissa kokonaan vai osittain ja miten pitkään. Selvitä, jos aineistoa tai sen osaa ei voida avata ja perustele se. Perusteluina voi olla luottamuksellisuus, liikesalaisuudet tai omistukselliset seikat (lisenssit, tekijänoikeus). Joskus aineistoa ei voida avata kohtuuttomien resurssivaatimusten takia (esim. aineisto vanhentuneessa formaatissa tai suuri analogisen aineiston määrä). Pohdi myös aineiston avaamisen mielekkyyttä: onko aineistollasi merkitystä muille, missä laajuudessa, onko aineistosi luvat kunnossa ja onko sinulla resurssia aineiston avaamiseen.

## Hyvät käytänteet

- Harkitse aineistosi jakamista jo tutkimuksen alussa
- Avoimuus ja aineiston jakaminen tehostaa sen uudelleen käyttöä
- Aineisto on suositeltavaa jakaa esim. Creative Commons –lisenssien avulla. Avoimille aineistoille suositellaan CC-BY lisenssiä
- On olemassa useita vaihtoehtoja pitkäaikaissäilytyksestä sekä jakamista varten
- Hyödynnä pysyviä tunnisteita mahdollistaaksesi aineistoon pääsyn pysyvän linkin avulla (esim. DOI, URN)

### Mallivastaus:

*Aineisto avataan ja siihen viitataan julkaisuissa. Muut tahot voivat hyödyntää aineistoa ottamalla yhteyttä tekijään. Data tallennetaan XML-formaattiin, joka mahdollistaa aineiston avoimen käytön.*

### Kysymys 9:

#### **Missä, miten ja kuka voi hyödyntää pitkäaikaissäilytettävää aineistoa?**

Pitkäaikaissäilytyksen tarkoituksena on tallentaa aineisto ja pitää se käyttökelpoisena sekä ymmärrettävänä kymmenien tai jopa satojen vuosien ajan. Aineistot, jotka valitaan pitkäaikaissäilytettäväksi, toimitetaan Karelian omaan, kansalliseen tai EU:n data-arkistoon. Pitkäaikaissäilytyksessä varmistetaan, että aineistot ovat löydettävissä, ymmärrettävissä, saatavissa ja käytettävissä myös tulevaisuudessa, sukupolvien ajan.

Kareliassa pitkäaikaissäilytyksen paikkoina ovat Karelia-PAS –verkkokansio (luna), mahdolliset erilliset pilvipalvelimet (erityisesti tutkimusympäristöjen suuret data-aineistot), sekä kansalliset ja kansainväliset palvelut kuten IDA, AILA ja EU Open Science Cloud. Kareliassa avattavien ja pitkäaikaissäilytettävien aineistojen metadata kuvataan Reportronic-järjestelmän Aineistot-osioon, josta on edelleen raajapinta Karelian www-sivulle (Avoimet Tutkimusaineistot). Merkittävien aineistojen metatiedot tallennetaan myös kansallisiin Etsin- /Qvain-palveluihin (Fairdata). Opinnäytetöiden metadatan keruu- ja tallentaminen ohjeistetaan erikseen.

## Hyvät käytänteet

- Kuvaile lyhyesti, mitkä aineistot on tarkoitus säilyttää ja kuinka pitkään niitä säilytetään. Kerro lisäksi, mitkä aineistot tuhoetaan projektin päätyttyä
- Muista selvittää rahoittaja- tai alakohtaiset sekä kansalliset suositukset liittyen aineiston säilytyspaikkoihin, data-arkistoihin tai tietokantoihin

### Mallivastaus:

*Hankkeen aineisto sijoitetaan uudelleen [tallennuspaikka], jotta varmistetaan, että tutkimusyhteisöllä on pitkäaikainen pääsy tutkimusaineistoon. Tallentamalla aineisto [tallennuspaikka], projektimme varmistaa sen, että aineisto tullaan siirtämään uusiin formaatteihin, alustoihin ja tallennusmedioihin hyvien käytänteiden mukaisesti. [Tallennuspaikka] generoi DOI :n (digital object identifier), jotta pääsy aineistoon on turvattu pysyvän linkin avulla.*

### Kysymys 10:

#### **Oletko huomionut aineistonhallinnan kustannukset?**

Aineiston laadukas hallinta ja avaaminen eivät synny ilman resursseja. Kustannuksia syntyy sekä asiantuntijatyöstä että mahdollisesti aineiston käsittelystä, tallentamisesta ja säilyttämisestä. Huomioi

hankkeen aikainen sekä hankkeen jälkeinen aineistohallinnan resursointi jo laatiessasi tutkimuksen budjettia.

### **Hyvät käytänteet**

- Muista huomioida aineistohallinnan kustannukset hankebudjetissa
- Tarvitsetko erikoisosaamista avataksesi ja tallentaaksesi aineiston? Esim. anonymisointiosaaminen.
- Onko tallennustilaa tarpeeksi vai tarvitsetko hankebudjettiin varoja lisäpalvelujen hankintaa varten?
- Mieti, tarvitsetko lisälaskentatehoa ja resursseja avataksesi aineiston ja mitä nämä kustannukset maksavat
- Miten aineistohallinnasta syntyneet kustannukset jaetaan partnereiden kesken yhteishankkeissa?

### **Mallivastaus:**

*Hankebudjetissa on allokoitu työaikaa ja kustannuksia kattamaan aineiston valmistelun ja dokumentoinnin arkistointia varten. [Tallennuspaikka] on arvioinut omien lisäkustannustensa olevan [euromäärä]. Tämä kustannus sisältyy myös hankebudjettiin*