

Aineistohallintasuunnitelman perusohje 2019

Miksi tutkimusaineistojen hallinta ja hallinnan suunnittelu on tärkeää?

- Tutkimusaineistojen hallinta ja aineistohallintasuunnitelman laatiminen ovat osa hyviä tieteellisiä käytäntöjä!
- Aineistojen katoamisen, häviämisen tai tuhoutumisen riski pienenee.
- Voit ennakoida omistajuuteen ja käyttöoikeuksiin liittyviä monimutkaisia ongelmia.
- Tuet avointa saatavuutta ja luot edellytyksiä tuleville menestyksekkäille yhteistyöhankkeille.
- Täytät rahoittajien vaatimukset.
- Säästät aikaa ja rahaa.
- Aineistohallintasuunnitelma kertoo osaamisestasi hankkeen johtajana.

Aineistohallintasuunnitelmassa aineisto (data) ymmärretään laajana käsitteenä. Englanninkielisen ohjeen data käännetään pääsääntöisesti suomenkielisessä ohjeessa aineistoksi. Aineistoa ovat ”kaikki tiedot, joita tarvitaan tutkimuksen toistamiseen, sekä kaikki muille mahdollisesti hyödyllinen materiaali”.
-Sarah Jones, Digital Curation Centre, UK.

Aineistohallintasuunnitelman tulisi kuvata, miten tutkimusaineistoja käsitellään **koko tutkimuksen elinkaaren ajan**. Aineistohallintasuunnitelmaa tulisi päivittää tutkimushankkeen edistyessä.

Aineistohallintakäytäntöjen tulisi perustua ns. **FAIR**-periaatteisiin, joiden mukaisesti aineiston tulee olla löydettävissä (Findable), saavutettavissa (Accessible), yhteentoimivaa (Interoperable) ja uudelleen käytettävissä (Re-usable).

Voit käyttää näitä yleisen tason kysymyksiä aineistohallintasuunnitelmasi jäsentelyyn. Kysymyksiin vastataan soveltuvin osin.

Onnea aineistohallintasuunnitelman laatimiseen!

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?

Ajattele aineistohallintasuunnitelmaa osana tutkimussuunnitelmaasi, jonka ei tarvitse olla luettavissa itsenäisenä asiakirjana. Aineistohallintasuunnitelma täydentää tutkimussuunnitelmaa kuvaamalla, miten aineistoa käsitellään teknisesti. Pällekkäisyyksien välttämiseksi voit viitata aineistohallintasuunnitelmassa tutkimussuunnitelmaan ja päinvastoin.

Kuvaa tässä lyhyesti, millaista aineistoa olet keräämässä tai tuottamassa. Selitä myös, millaista olemassa olevaa aineistoa aiot käyttää. (Esim. tekstejä, kuvia, valokuvia, mittaustuloksia, tilastoja, fyysisiä näyttöitä tai koodia).

Luokittele aineistosi siten, että voit viitata siihen jäljempänä suunnitelmassa. Vastauksesi tähän ensimmäiseen kysymykseen voi siis muodostaa suunnitelman loppuosan perusrakenteen - esimerkiksi: A) tätä hanketta varten kerätty aineisto, B) prosessin tuloksena syntynyt aineisto, C) aiemmin kerätty aineisto,

jota käytetään uudelleen tässä hankkeessa, D) hankkeen hallinnointiin liittyvät asiakirjat ja hankkeen raportit.

Mainitse kunkin tietoaineiston tiedostomuoto. Joissain tapauksissa tutkimushankkeen aikana käytetyt tiedostomuodot voivat erota aineiston arkistoinnissa käytetyistä tiedostomuodoista. Mainitse tällöin molemmat tiedostomuodot. Tiedostomuoto vaikuttaa oleellisesti aineiston saavutettavuuteen ja uudelleenkäytettävyyteen.

Vinkejä hyvistä käytännöistä

- Tutkimussuunnitelmassa tulee kuvata seikat, jotka liittyvät aineistoon sisältyvien tietojen analysointiin ja käytettyihin menetelmiin.
- Tiedostomuotoja ovat esimerkiksi .csv, .txt, .docx, .xlsx ja .tif.
- Kun kerrot käyttämäsi tiedostomuodot, muista mainita tiedostojen tarkasteluun tai käyttämiseen tarvittavat erityiset tai epätavalliset ohjelmat - varsinkin, jos ne ohjelmoidaan hankkeessasi.
- Käytä taulukkoa tai luetteloa esittämään datatyytit, tiedostomuodot, käytettävät ohjelmat ja muut seikat tiiviissä muodossa.
- Voit myös arvioida aineiston määrän kasvua, esim. viikkotasolla: "Tutkimusprojektin tuottaa/kerää aineistoa noin 100 Gt:a viikossa."

1.2 Miten aineiston yhtenäisyys ja laatu varmistetaan?

Selitä, miten aineiston keruu-, analysointi- ja käsittelymenetelmät voivat vaikuttaa aineiston laatuun ja miten minimoit aineiston tarkkuuteen ja virheettömyyteen liittyvät riskit.

Aineiston laadunhallinnalla varmistetaan, että aineisto ei epähuomiossa muutu ja että aineisto pysyy virheettömänä koko sen elinkaaren ajan. Laatuongelmia voi syntyä aineistojen teknisen käsittelyn, konvertoinnin tai siirron vuoksi tai sisällön käsittelyn ja analyysin aikana.

Vinkejä hyvistä käytännöistä

- Äänitettyjen tai kuvattujen haastattelujen litteroinnit tulisi tarkistaa myös jonkun muun kuin alkuperäisen litteroijan.
- Analogisen aineiston digitoinnissa tulisi pyrkiä tarkkuuden vuoksi mahdollisimman korkeaan resoluutioon.
- Tietoja muunnettaessa alkuperäisen sisällön säilyminen tulisi aina varmistaa.
- Käytettävien ohjelmistojen tulisi laskea tarkistussummia.

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)?

Kerro, miten täytät tiukat eettiset vaatimukset ja noudatat asiaankuuluvaa lainsäädäntöä tutkimusaineiston hallinnassa. Eettiset seikat tulee ottaa huomioon koko tutkimusaineiston elinkaaren ajan.

Useimpia tutkimushankkeita koskee esimerkiksi eettinen vaatimus tutkittavien informointiin liittyvien ohjeiden noudattamisesta. Jos käsittelet henkilökohtaisia tai arkaluonteisia tietoja, kerro, miten suojaat tutkittavien yksityisyyden ja anonymisoit tai pseudonymisoit tiedot.

Vinkejä hyvistä käytännöistä

- Tutustu organisaatiosi eettisiin ohjeisiin ja tietoturvaperiaatteisiin. Valmistaudu noudattamaan kyseisten asiakirjojen ohjeistusta.

- Jos eettinen toimikunta antaa lausunnon tutkimuksestasi, kerro aineistohallintasuunnitelmassa, miten noudatat toimintaohjeita (esim. miten henkilökohtaiset tai arkaluonteiset tiedot poistetaan aineistosta ennen sen jakamista, jotta yksityisyydensuoja voidaan varmistaa).
- Lisätietoja hyvästä tieteellisestä käytännöstä on esimerkiksi Tutkimuseettisen neuvottelukunnan [verkkosivuilla](#).
- Tutustu myös esimerkiksi eurooppalaiseen tutkimuseettiseen toimintaohjeistoon ([European Code of Conduct for Research Integrity](#)).

2.2 Miten aineiston omistajuuteen, tekijänoikeuksiin ja immateriaalioikeuksiin liittyviä asioita hallitaan? Estävätkö tekijänoikeudet, käyttöoikeudet tai muut rajoitukset aineiston käyttämisen tai jakamisen?

Kerro, kuka aineiston omistaa ja mitä omistajuudesta on sovittu. Kerro myös, kuka voi antaa luvan aineiston (uudelleen) käyttämiseen.

Vinkejä hyvistä käytännöistä

- Tarkista tietojen omistajuutta ja käyttö- ja jakeluoikeuksia koskevat organisaatiosi periaatteet.
- Omistajuutta koskevat sopimukset tulisi tehdä mahdollisimman varhaisessa vaiheessa hankkeen elinkaarta.
- Ota huomioon myös rahoittajan periaatteet tekijän- ja immateriaalioikeuksiin liittyen.
- Tutkimusaineiston sekä kaikkien tutkimushankkeessa luotujen koodien ja ohjelmistojen tarjoaminen uudelleen käytettäväksi on suositeltavaa. Käytä tähän tarkoitukseen esimerkiksi [Creative Commons](#)-, [GNU](#)- tai [MIT](#)-lisenssiä tai muuta asianmukaista lisenssiä.

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?

Aineistoa koskeva dokumentaatio mahdollistaa sen, että muut käyttäjät (sekä ihmiset että tietokoneet) voivat löytää ja käyttää tietoaineistoja ja tiedostoja sekä viitata niihin asianmukaisesti. Aineiston metatiedot ovat oleellisia, jotta voidaan kertoa, missä, milloin, miksi ja miten aineisto kerättiin, käsiteltiin ja tulkittiin. Metatiedot voivat sisältää kuvailua myös koeasetelmasta, analyysimenetelmästä ja tutkimusympäristöstä.

Vinkejä hyvistä käytännöistä

- Kuvaa tarjoamasi dokumentaatio (esim. README-tiedostot ja metatiedot), jotka auttavat jatkokäyttäjiä löytämään, ymmärtämään ja uudelleen käyttämään aineistoa.
- Ns. [FAIR](#)-periaatteiden noudattaminen auttaa varmistamaan, että aineisto on löydettävissä (Findable), saavutettavissa (Accessible), yhteentoimivaa (Interoperable) ja uudelleen käytettävissä (Re-usable).
- Käytä tutkimusvälineitä, joiden luomat metatiedot ovat automaattisesti standardien mukaisia. Tällöin aineistosi on siirrettävissä eri valmistajien työkalusta toiseen.
- Pohdi, miten aineisto organisoidaan hankkeen aikana. Kuvaa esimerkiksi tiedostojen nimeämistä koskevat käytännöt, versionhallinta ja kansiorakenne.
- Erittele tarvittavat tiedot, jotka tulisi tallentaa, jotta myös muut tutkijat voivat löytää, tarkastella, tulkita ja käyttää aineistoasi sekä viitata siihen.
- Arkistot edellyttävät usein tietyn metatietostandardin käyttämistä. Tarkista, onko käytettävissä ala-, yhteisö- tai arkistokohtaista metatietomallia tai -standardia.

4. Tallentaminen ja varmuuskopiointi tutkimushankkeen aikana

4.1 Minne aineistosi tallennetaan ja miten se varmuuskopioidaan?

Kuvaa, minne aineisto tallennetaan ja varmuuskopioidaan tutkimushankkeen aikana. Aineiston säilyttämis- ja jakamismenetelmät tutkimushankkeen päätyttyä kuvataan yksityiskohtaisesti kohdassa 5.

Pohdi, kuka vastaa tietojen varmuuskopioinnista ja palauttamisesta. Jos mukana on useita tutkijoita, tee suunnitelma yhteistyökumppaniesi kanssa ja varmista tietojen turvallinen siirto osallistujien välillä.

Vinkejä hyvistä käytännöistä

- On suositeltavaa, että aineisto tallennetaan oman organisaatiosi tarjoamaan ja ylläpitämään turvalliseen paikkaan.

4.2 Kuka valvoo pääsyä aineistoon ja miten suojattua pääsyä aineistoon valvotaan?

Tietoturvasta huolehtiminen on oleellista varsinkin, jos aineisto on arkaluonteista (esim. sisältää henkilökohtaisia tai poliittisesti arkaluonteisia tietoja tai liikesalaisuuksia). Kerro, kenellä on pääsy aineistoosi, mitä kyseiset henkilöt voivat tehdä aineistolle ja miten varmistat aineiston turvallisen siirron yhteistyökumppaneillesi.

Vinkejä hyvistä käytännöistä

- Aineistoon pääsynhallinnan toteutus riippuu siitä kuinka arkaluontoinen aineisto kulloinkin on kyseessä.

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?

Kuvaa, julkaisetko aineiston tai asetatko sen muuten avoimesti saataville joko kokonaan tai osittain. Jos aineistoasi ei voida avata kokonaan tai osittain, perustele miksi. Tutkimusaineiston avoimuus edistää aineiston uudelleenkäyttöä.

Vinkejä hyvistä käytännöistä

- Voit julkaista aineistosi kuvauksen (ts. metatiedot) asettamatta aineistoa avoimesti saataville ja näin rajoittaa pääsyä aineistoon, vaikka tieto aineistosi olemassaolosta on julkinen.
- Julkaise aineistosi asianmukaisessa arkistossa tai vertaisarvioidussa lehdessä.
- Voit etsiä sopivaa arkistoa osoitteesta re3data.org.
- Muista tarkistaa rahoittajan omat, alakohtaiset tai kansalliset arkistosuositukset.
- Tutkimusaineiston sekä kaikkien tutkimushankkeessa luotujen koodien ja ohjelmistojen tarjoaminen uudelleen käytettäväksi on suositeltavaa. Käytä tähän tarkoitukseen esimerkiksi [Creative Commons](https://creativecommons.org/)-, [GNU](https://www.gnu.org/licenses/)- tai [MIT](https://www.mit.edu/~dts/opensource/)-lisenssiä tai muuta asianmukaista lisenssiä.
- Pyri käyttämään sellaista arkistoa tai julkaisijaa, joka tarjoaa aineistollesi pysyvän tunnusteen (esim. DOI, URN)

5.2 Mihin pitkällä aikavälillä arvokkaat tiedot arkistoidaan ja kuinka pitkäksi ajaksi?

Kuvaa lyhyesti, mitkä tiedot arkistoidaan ja kuinka pitkäksi ajaksi sekä mitkä tiedot hävitetään hankkeen päätyttyä. Kuvaa arkistoitujen tietojen saatavuutta koskevat periaatteet.

Vinkejä hyvistä käytännöistä

- Muista tarkistaa rahoittajan omat, alakohtaiset tai kansalliset arkistosuositukset.

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).

Tyypillisesti osa vastuihin liittyvistä yksityiskohdista on kuvattu jo aiempien kysymysten kohdalla. Voit vaihtoehtoisesti joko viitata aiempiin kysymyksiin tai koostaa tähän kaikki vastuutahot sekä niiden vaatimat resurssit.

Arvioi, edellyttääkö aineiston hallinnointi, säilyttäminen ja jakaminen asiantuntija-apua. Pohdi, mitä täydentäviä tietoteknisiä välineitä, palveluita ja resursseja tarvitaan ja mitkä niiden kustannukset ovat.

Vinkkejä hyvistä käytännöistä

- Erittele aineistonhallintakustannukset budjetissa rahoittajan ohjeiden mukaisesti.