

Kansalliset DMP (aineistohallintasuunnitelma) -ohjeet 2020

Miten laadin aineistohallintasuunnitelma? LUE AINAKIN NÄMÄ ENSIN.

- Tutustu ensin kaikkiin kysymyksiin!
- Käytä aineistohallintasuunnitelmaa tutkimussuunnitelmasi täydentämiseen. Vältä päällekkäisyyttä tutkimussuunnitelman kanssa.
 - Tutkimussuunnitelma kuvaa aineiston tieteellistä, analyttistä ja metodologista käsittelyä.
 - Aineistohallintasuunnitelma kuvaa aineiston teknistä ja hallinnollista käsittelyä.
 - Päällekkäisyyksien välttämiseksi viittaa aineistohallintasuunnitelmassa tutkimussuunnitelmaan ja päinvastoin.
- Käytä aineistohallintasuunnitelmaa riskinarvioinnin kuvauksena. Osoita, että pystyt tunnistamaan, ennakoimaan ja hallitsemaan aineistohallintaprosessiin liittyviä riskejä.
- Laadi aineistohallintasuunnitelma oman tutkimushankkeesi näkökulmasta. Älä kopioi esimerkkejä muualta.
- Kirjoita vain virkeitä, jotka itse ymmärrät.
- Vastaa kysymyksiin, jotka koskevat tutkimustasi. Jos jokin kysymys ei liity tutkimukseesi, perustele, miksi näin on.
- Vastaa ainakin pääkysymyksiin. Kaikkiin alakysymyksiin ei tarvitse vastata erikseen.
- Sisällytä suunnitelmaan taustatietoja, esimerkiksi hakijan ja hankkeen nimet, hankkeen numero, rahoituspäätöksen tunniste ja aineistohallintasuunnitelman versiotieto.
- Osoita aineiston- ja versionhallintataitosi esimerkiksi nimeämällä myös aineistohallintasuunnitelma asianmukaisesti.
- Noudata organisaation tai rahoittajan vaatimuksia.

Mihin tutkimusaineistojen hallinnointia ja aineistohallintasuunnitelmaa tarvitaan?

- Tutkimusaineistojen hallinnointi ja aineistohallintasuunnitelman laatiminen kuuluvat hyviin tieteellisiin käytäntöihin.
- Etukäteen laadittu aineistohallintasuunnitelma vähentää aineistojen häviämisen tai tuhoutumisen riskiä.
- Suunnitelman avulla pystyt ennakoimaan ja hallitsemaan omistajuuteen ja käyttöoikeuksiin liittyviä yksityiskohtia.
- Etukäteen laadittu suunnitelma mahdollistaa aineistojen avaamisen ja luo edellytyksiä uusien yhteistyöhankkeiden muodostumiselle.
- Suunnittelu auttaa seuraamaan rahoittajien linjauksia.
- Suunnittelemalla säästät aikaa ja rahaa.
- Aineistohallintasuunnitelma kertoo osaamisestasi hankkeen johtajana.

Aineistohallintasuunnitelman yhteydessä aineisto (data) ymmärretään laajasti. Tutkimusaineisto ja -data kattavat kaikki tiedot ja resurssit, joihin tutkimustuloksesi perustuvat. Suunnitelmassa voit keskittyä kuvamaan vastuullasi olevaan aineistoon.

Aineistohallintasuunnitelma kuvaa, miten tutkimusaineistoa hallinnoidaan koko tutkimuksen elinkaaren ajan. Aineistohallintasuunnitelmaa päivitetään säännöllisesti tutkimushankkeen edistessä.

Tutkimusaineiston hallintaan liittyvien käytäntöjen tulee pyrkiä ns. FAIR-periaatteiden noudattamiseen eli siihen, että aineisto on löydettävissä (Findable), saavutettavissa (Accessible), yhteentoimiva (Interoperable) ja uudelleen käytettävissä (Re-usable).

Onnea aineistohallintasuunnitelman laatimiseen!

Kansalliset DMP (aineistohallintasuunnitelma) -ohjeet 2020

Tuuli-projekti: <https://wiki.helsinki.fi/x/b3eZCO>

Päivitetty 24.1.2020



1. Aineiston yleiskuvaus

1.1 Millaiseen aineistoon tutkimuksesi perustuu? Millaista aineistoa kerätään, tuotetaan tai käytetään uudelleen? Missä tiedostomuodossa aineisto on? Esitä myös karkea arvio tuotettavan/kerättävän aineiston koosta.

Kuvaa lyhyesti, millaista aineistoa olet keräämässä tai tuottamassa. Selitä, millaista jo olemassa olevaa aineistoa aiot (uudelleen)käyttää. Luettelo esimerkiksi tekstityypit, kuvat, valokuvat, mittaukset, tilastot, fyysiset näytteet tai koodit.

Luokittele aineistosi taulukossa tai selkeällä luettelolla esimerkiksi seuraavasti:

- A) tätä hanketta varten kerätty aineisto
- B) tutkimusprojektissa tuotettu aineisto
- C) aiemmin kerätty aineisto, jota käytetään tässä hankkeessa uudelleen
- D) tutkimuksen hallintaan liittyvät asiakirjat.

Luokittelu kannattaa muotoilla tutkimuksessa hyödyntämäsi ja tuottamasi aineiston käyttöoikeuksien ja lisenssien perusteella. Kerro lyhyesti, minkä lisenssin perusteella sinulla on oikeus aineiston (uudelleen)käyttämiseen. Kohdassa 1.1 muotoilemasi luokittelu muodostaa luontevan jäsentelyn, johon voit viitata aineistohallintasuunnitelman muissa kysymyksissä.

Kuvaa käyttämäsi tiedostomuodot. Joissain tapauksissa tutkimushankkeen aikana käytettävät tiedostomuodot voivat erota tiedostomuodoista, joita käytetään hankkeen jälkeen aineiston arkistoinnissa. Mainitse tällöin molemmat tiedostomuodot. Tiedostomuoto vaikuttaa oleellisesti aineiston saavutettavuuteen ja uudelleenkäytettävyyteen.

Sisällytä aineistohallintasuunnitelmaan kuvaus tarvittavasta levytilasta, ei siitä, kuinka monta tiedonantajaa hankkeeseen osallistui. Aineiston koon arviointi summittaisesti riittää (esim. alle 100 Gt, noin 1 Tt tai useita petatavuja).

Vinkkejä hyvistä käytännöistä

- Käytä taulukkoa tai luetteloa esittämään tietotyypit, tiedostomuodot, käytettävät ohjelmat ja aineiston koko tiiviissä muodossa.
- Mahdollisia tiedostomuotoja ovat .csv, .txt, .docx, .xlsx ja .tif.
- Muista mainita tietojen tarkasteluun tai käyttämiseen tarvittavat erityiset tai epätavalliset ohjelmat varsinkin, jos ne koodataan tai tuotetaan hankkeessasi.
- Voit myös arvioida, kuinka paljon hankkeessa tuotetaan tai kerätään aineistoa esimerkiksi viikossa: "Hankkeessa tuotetaan/kerätään viikoittain noin 100 Gt aineistoa."
- **VÄLTÄ PÄÄLLEKKÄISYYTTÄ TUTKIMUSSUUNNITELMAN KANSSA!** Aineiston analysointiin ja tutkimusmenetelmiin liittyvät yksityiskohdat ja prosessit kuvataan tutkimussuunnitelmassa.

1.2 Miten aineiston yhtenäisyys ja laatu varmistetaan?

Selitä, miten aineiston keruun, analysoinnin ja käsittelyn menetelmät voivat vaikuttaa aineiston laatuun, ja miten turvaat aineiston virheettömyyden.

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

Vinkkejä hyvistä käytännöistä

- On suositeltavaa, että haastattelujen ääni- tai videotalltointien litteroinnit tarkistaa joku muu kuin litteroija.
- Virheiden ja epätarkkuuksien välttämiseksi analoginen aineisto on syytä digitoida mahdollisimman korkealla resoluutiolla.
- Alkuperäisen tietosisällön säilyminen on aina varmistettava tietoja ja tiedostoja muunnettaessa.
- Käytä ohjelmistoja, jotka laskevat tarkistussummia.
- Tarjoa koulutusta ja ohjeita varmistaaksesi, että kaikki tutkimusryhmäsi jäsenet pystyvät huolehtimaan laadunvalvonnasta ja ennakoimaan laatuun liittyvät riskit.
- VÄLTÄ PÄÄLLEKKÄISYYTTÄ TUTKIMUSSUUNNITELMAN KANSSA! Aineiston analysointiin, menetelmiin ja työvälineisiin liittyvät seikat kuvataan tutkimussuunnitelmassa. Aineistohallintasuunnitelmassa ei siis kuvata esimerkiksi tutkimuslaitteiden kalibrointia.

2. Eettisten periaatteiden ja lainsäädännön noudattaminen

2.1 Mitä juridisia seikkoja liittyy aineiston hallintaan (esim. EU:n yleinen tietosuojasetus ja muu aineiston käsittelyyn liittyvä lainsäädäntö)?

Jokaiseen tutkimusaineistoon liittyy oikeuksia, lainsäädäntöä ja tutkimusetiikkaa koskevia kysymyksiä. Osoita, että tunnet aineiston käsittelyyn liittyvän lainsäädännön. Jos käsittelet henkilötietoja tai arkaluonteisia tietoja, kerro, miten suojaat tutkittavien yksityisyyden ja anonymisoi tai pseudonymisoi aineiston.

Vinkkejä hyvistä käytännöistä

- Tutustu organisaatiosi eettisiin ohjeisiin, yksityisyyttä koskeviin ohjeisiin ja tietoturvaperiaatteisiin ja valmistaudu noudattamaan kyseisten asiakirjojen ohjeistusta.
- Jos eettinen toimikunta antaa lausunnon tutkimuksestasi, kerro aineistohallintasuunnitelmassa, miten noudatat toimintaohjeita (ts. miten henkilökohtaiset tai arkaluonteiset tiedot poistetaan aineistosta ennen sen jakamista, jotta yksityisyydensuoja voidaan varmistaa).
- Jos käsittelet henkilötietoja, yksilöi, millaisia henkilötietoja keräät.
- Kaikki tunnistettua tai tunnistettavaa henkilöä koskevat tiedot ovat henkilötietoja. Myös nimet, puhelinnumerot, sijaintitiedot ja henkilön isovanhempien synnynnäisten sairauksien kaltaiset tiedot ovat henkilötietoja.

- Lisätietoja ks. mm. Tietosuojavaltuutetun toimisto (<https://tietosuoja.fi/henkilotietojen-kasittely>)
- VÄLTÄ PÄÄLLEKKÄISYYTTÄ TUTKIMUSSUUNNITELMAN KANSSA! Yksityiskohtaiset tutkimuseettiset seikat, eettisten toimikuntien lausunnot ja koe-eläinten käyttö ym. kuvataan tutkimussuunnitelmassa.

2.2 Miten hallinnoit käyttämäsi, tuottamasi ja jakamasi aineiston oikeuksia?

Kuvaa, miten sovit hankkeessa kerätyn, tuotetun ja (uudelleen)käytetyn tutkimusaineiston käyttöoikeuksista. Voit käyttää luokittelua, joka esitetään edellä kohdassa 1. Kuhunkin kysymyksessä 1.1 luokiteltuun ryhmään liittyy erilaisia oikeuksia ja lisenssejä. Kuvaa oikeuksien siirtämiseen liittyvät menettelyt, jotka koskevat hankettasi. Kuvaa hankettasi koskevat salassapitokäytännöt.

Vinkkejä hyvistä käytännöistä

- Tarkista tietojen omistajuutta ja käyttö- ja jakeluoikeuksia koskevat organisaatiosi periaatteet.
- Tarkista, että olet saanut luvat aineiston säilyttämiseen ja jakamiseen.
- Omistajuutta ja käyttöoikeuksia koskevat sopimukset tehdään mahdollisimman varhaisessa vaiheessa hankkeen elinkaarta.
- Huomioi rahoittajan linjaukset.
- Tutkimusaineiston sekä kaikkien tutkimushankkeessa luotujen koodien ja ohjelmistojen tarjoaminen uudelleen käytettäväksi on suositeltavaa. Käytä tähän tarkoitukseen esimerkiksi Creative Commons - (<https://creativecommons.org/choose/>), GNU- (<https://www.gnu.org/licenses/gpl-3.0.en.html>) tai MIT-lisenssiä (<https://opensource.org/licenses/MIT>) tai muuta asianmukaista lisenssiä.

3. Dokumentointi ja metatiedot

3.1 Miten dokumentoit aineistosi, jotta se on löydettävissä, saavutettavissa, yhteentoimiva 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. Dokumentaatio sisältää oleellisia tietoja esimerkiksi siitä, missä, milloin, miksi ja miten aineisto kerättiin, käsiteltiin ja tulkittiin. Ilman asianmukaista dokumentaatiota aineisto on hyödytön. Kerro, mitä työkalua (esim. Qvain) aiot käyttää aineistosi kuvailuun. Älä mainitse metatietostandardeja, jos et aio käyttää niitä. Voit ennakoida hankkeesi kuvailutietojen avaamista jo tässä kohdassa. Kerro avaamisesta yksityiskohtaisesti alla kohdassa 5.

VÄLTÄ PÄÄLLEKKÄISYYTTÄ TUTKIMUSSUUNNITELMAN KANSSA!

Aineistotason dokumentaatiota (<https://www.ukdataservice.ac.uk/manage-data/document/data-level.aspx>) ja kokeita, analyysimenetelmiä ja tutkimuskontekstia koskevat tiedot kuuluvat tutkimussuunnitelmaan.

Aineistonhallintasuunnitelmassa voit keskittyä tutkimustason dokumentaatioon (<https://www.ukdataservice.ac.uk/manage-data/document/study-level.aspx>).

Vinkkejä hyvistä käytännöistä

- Kuvaa kaikki tarjoamasi dokumentaatiotyypit (esim. README-tiedostot ja metatiedot), jotka auttavat jatkokäyttäjiä löytämään, ymmärtämään ja uudelleen käyttämään aineistoa.
- Ns. FAIR-periaatteiden (<https://www.fairdata.fi/miksi-fairdata/fair-periaatteet/>) noudattaminen auttaa varmistamaan, että aineisto on löydettävissä, saavutettavissa, yhteentoimiva ja uudelleen käytettävissä.
- Lisätietoja aineiston dokumentaation vähimmäisvaatimuksista on esimerkiksi täällä: (<https://www.fairdata.fi/qvain/qvain-light-kayttoopas/>).
- Käytä välineitä ja laitteita, jotka tuottavat standardoitua metadataa.
- Tunnista ne vähimmäistason tiedot, jotka tulee tallentaa, jotta muut tutkijat voivat löytää, tarkastella, tulkita ja käyttää aineistoasi sekä viitata siihen.

4. Tallentaminen ja varmuuskopiointi tutkimushankkeen aikana

4.1 Minne aineistosi tallennetaan, ja miten se varmuuskopioidaan?

Kuvaa, minne aineisto tallennetaan ja varmuuskopioidaan tutkimushankkeen aikana. Kuvaa kohdassa 5 yksityiskohtaisesti menetelmät, joita käytetään aineiston säilyttämiseen ja jakamiseen tutkimushankkeen päätyttyä.

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

Osoita, että tunnet organisaatiosi säilytysratkaisut. Älä viittaa ainoastaan tietotekniikkapalveluihin. Perimmäinen vastuu aineistosi käsittelemisestä on sinulla, ei tietotekniikkaosastolla tai organisaatiolla.

Vinkkejä hyvistä käytännöistä

- On suositeltavaa, että aineisto tallennetaan oman organisaatiosi tarjoamaan ja ylläpitämään turvalliseen paikkaan.
- **ÄLÄ KÄYTÄ** ulkoisia kovalevyjä pääasiallisena säilytysratkaisuna.

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

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

Vinkkejä hyvistä käytännöistä

- Aineistoon pääsy tulee aina valvoa luottamuksellisuustaso mukaan.

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?

Kerro, julkaisetko tai asetatko aineistosi saataville kokonaisuudessaan vai vain osittain. Jos aineistoasi ei voida avata kokonaan tai osittain, perustele se.

Jos kyse on arkaluonteisista tiedoista, joita ei voi avata, kuvaa niihin liittyvien metatietojen avaaminen. Kuvaa arkaluonteisen aineiston turvallinen säilytysmenettely seuraavassa kohdassa 5.2.

Tutkimusaineiston avoimuus edistää aineiston uudelleenkäyttöä.

Vinkkejä hyvistä käytännöistä

- Voit julkaista aineistosi kuvauksen (ts. metatiedot) asettamatta aineistoa avoimesti saataville ja näin rajoittaa pääsyä aineistoon.
- Julkaise aineistosi data-arkistossa (data repository) tai datajulkaisuissa (data journal).
- Voit etsiä sopivaa data-arkistoa osoitteesta re3data.org (<https://www.re3data.org/>).
- Muista tarkistaa rahoittajan omat, alakohtaiset tai kansalliset suositukset data-arkiston valintaan liittyen.
- Tutkimusaineiston sekä kaikkien tutkimushankkeessa luotujen koodien ja ohjelmistojen tarjoaminen uudelleen käytettäväksi on suositeltavaa. Käytä tähän tarkoitukseen esimerkiksi Creative Commons - (<https://creativecommons.org/choose/>), GNU- (<https://www.gnu.org/licenses/gpl-3.0.en.html>) tai MIT-lisenssiä (<https://opensource.org/licenses/MIT>) tai muuta asianmukaista lisenssiä.
- Harkitse sellaisten arkistojen tai julkaisijoiden käyttämistä, jotka tarjoavat pysyvän tunnusteen (PID) ja varmistavat näin, että aineisto on käytettävissä pysyvän linkin (esim. DOI, URN) kautta.
- VÄLTÄ PÄÄLLEKKÄISYYTTÄ JULKAISUSUUNNITELMAN KANSSA! Tutkimusartikkelin julkaiseminen ei ole sama asia kuin aineiston julkaiseminen. Datajulkaisut (data journal) ovat tutkimusaineistojen julkaisuun erikoistuneita foorumeita.

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

Kuvaa lyhyesti, mitkä osat aineistostasi aiot säilyttää ja kuinka pitkään. Luokittele aineistosi ennakkoidun säilytysajan mukaan:

- A) Hankkeen päättyessä tuhottava aineisto
- B) Aineisto, joka arkistoidaan verifiointin ajaksi (tieteenalan mukaan esim. 5–15 vuotta)
- C) Mahdollista uudelleenkäyttöä varten arkistoitava aineisto (esim. 25 vuotta)
- D) Pitkällä aikavälillä arvokas aineisto, joka arkistoidaan kuratoidussa ympäristössä kymmeniksi tai sadoiksi vuosiksi tulevia sukupolvia varten

Kuvaa, mitkä osat aineistostasi hävitetään hankkeen jälkeen ja miten. Kuvaa arkistoitujen tietojen saatavuutta koskevat periaatteet. Harkitse kuratointiperiaatteita noudattavien arkistojen käyttöä.

Vinkkejä hyvistä käytännöistä

- Muista tarkistaa rahoittajan omat, alakohtaiset tai kansalliset tietoarkistosuosituksen.

6. Aineistonhallintaa koskevat vastuut ja resurssit

6.1 Kuka (esim. tehtävä ja laitos) vastaa aineistonhallinnasta (ns. tietovastaava)?

Kerro tässä lyhyesti, miten edellisissä vastauksissa kuvatut tehtävät ja vastuut on jaettu

Vinkkejä hyvistä käytännöistä

- Hahmottele aineistonhallintaan/tietovastaavan toimintaan liittyvät tehtävät ja vastuut, jotka liittyvät esimerkiksi tiedonkeruuseen, metatietojen tuottamiseen, aineiston laatuun, säilytykseen ja varmuuskopiointiin sekä aineiston arkistointiin ja jakamiseen. Jos mahdollista, nimeä vastuuhenkilö(t).
- Yhteistyöhankkeissa kuvaa kumppaneiden aineistonhallintavastuut.
- Määrittele, kuka vastaa aineistonhallintasuunnitelman toteuttamisesta ja varmistaa, että suunnitelma tarkastetaan ja tarvittaessa korjataan.
- Harkitse aineistonhallintasuunnitelman säännöllistä päivittämistä.
- Pohdi myös, kuka vastaa aineistosta, jota hankkeesi tuottaa vielä päätyttyään.

6.2 Mitä resursseja aineistonhallinta edellyttää, jotta voit varmistaa, että aineisto voidaan avata ja säilyttää FAIR-periaatteiden (ks. edellä) mukaan?

Arvioi resurssit (esim. raha ja aika), joita tarvitaan aineiston hallintaan, säilyttämiseen ja jakamiseen. Pohdi, mitä täydentäviä tietoteknisiä tiloja ja resursseja tarvitaan ja mitkä niiden kustannukset ovat.

Vinkkejä hyvistä käytännöistä

- Erittele aineistonhallintakustannukset budjetissa rahoittajan vaatimusten mukaisesti.

- Harkitse tarkkaan, mitä resursseja tarvitaan aineiston jakamiseen, säilyttämiseen ja kuratointiin. Perustele niiden tarpeellisuus. Tarvittavia resursseja ovat esimerkiksi säilytyskustannukset, laitteet, työaika sekä mahdolliset data-arkiston tallennuskustannukset. Huomio, että myös aineiston saataville asettamisen valmistelusta aiheutuu kustannuksia ja se vie aikaa. Kerro, miten kaikki yllä mainitut on otettu huomioon kustannuslaskelmassa.