

# Kansalliset DMP (aineistonhallintasuunnitelma) -ohjeet 2021

Aineistonhallintasuunnitelmasi (DMP) on elävä dokumentti, jossa kuvaillet, kuinka hallitset tutkimusaineistosi ja -datasi koko tutkimuksen elinkaaren ajan. Päivitä suunnitelma projektisi edetessä. Päällekkäisyyksien välttämiseksi voit viitata aineistonhallintasuunnitelmassasi tutkimussuunnitelmaasi ja päinvastoin. Tutkimussuunnitelma kuvaa aineiston ja datan tieteellistä, analyttistä ja metodologista käsittelyä, kun taas DMP kuvaa niiden teknistä ja hallinnollista käsittelyä. Aineistonhallintasuunnitelman yhteydessä aineisto (*data*) ymmärretään laajasti. Tutkimusaineisto ja -data kattavat kaikki tiedot ja resurssit, joihin tutkimustuloksesi perustuvat.

## Miten laadin aineistonhallintasuunnitelma? LUE AINAKIN NÄMÄ ENSIN.

- Tutustu ensin kaikkiin kysymyksiin! Vastaa kysymyksiin soveltuvin osin ja ainakin pääryhmiin - jos tietty kysymys ei ole sovellettavissa tapauksessasi, perustele miksi näin.
- Sisällytä suunnitelmasi alkuun taustatiedot, kuten hakijan ja projektin nimi, projektin numero, rahoitusohjelma ja DMP-versio. Osoita aineiston- ja versionhallintataitosi nimeämällä myös aineistonhallintasuunnitelmasi asianmukaisesti.
- Käytä aineistonhallintasuunnitelmaa riskinarvioinnin kuvauksena. Osoita, että pystyt tunnistamaan, ennakoimaan ja hallitsemaan aineistonhallintaprosessiin liittyviä riskejä.
- Laadi aineistonhallintasuunnitelma oman tutkimushankkeesi näkökulmasta. Älä kopioi esimerkkejä muualta.
- Noudata organisaation tai rahoittajan vaatimuksia.

## Mihin tutkimusaineistojen hallinnointia ja aineistonhallintasuunnitelmaa tarvitaan?

- Tutkimusaineistojen hallinta ja aineistonhallintasuunnitelman laatiminen kuuluvat hyviin tieteellisiin käytäntöihin. Ne auttavat sinua säästämään aikaa ja rahaa!
- Etukäteen laadittu aineistonhallintasuunnitelma vähentää aineistojen häviämisen tai tuhoutumisen riskiä.
- Suunnitelman avulla pystyt ennakoimaan ja hallitsemaan omistajuuteen ja käyttöoikeuksiin liittyviä yksityiskohtia.
- DMP:n kirjoittaminen ohjaa noudattamaan FAIR-periaatteita (Findable, Accessible, Interoperable ja Re-usable), mikä lisää tutkimusaineistojen ja -datan uudelleenkäyttöä sekä lisää projektisi näkyvyyttä.
- Suunnittelu auttaa seuraamaan rahoittajien linjauksia.
- Aineistonhallintasuunnitelma kertoo osaamisestasi hankkeen johtajana.

## 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) aiemmin kerätty aineisto, jota käytetään tässä hankkeessa uudelleen,
- B) tätä hanketta varten kerätty aineisto,
- C) tutkimusprojektissa tuotettu aineisto.

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

- Hyödynnä jo käytössä olevia versionhallintaprosesseja - esimerkiksi yleisesti käytössä olevia tiedostojen nimeämiskäytäntöjä tai Git-alustojen tarjoamia työkaluja.
- On suositeltavaa, että haastattelujen ääni- tai videotaltiointien 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.
- 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. Aineistonhallintasuunnitelmassa ei siis kuvata esimerkiksi tutkimuslaitteiden kalibrointia.

## 2. Eettisten periaatteiden ja lainsäädännön noudattaminen

### 2.1 Mitä juridisia seikkoja liittyy aineiston hallintaan (esim. EU:n yleinen tietosuoja-asetus 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 aineistonhallintasuunnitelmassa, 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. Ilman asianmukaista dokumentaatiota aineiston jatkokäyttö on mahdotonta.

Aineiston ja datan [dokumentaatio](#) sisältää olennaisia tietoja, esimerkiksi a) **metadata** (löytämistä ja tunnistamista varten) missä, milloin, miksi ja miten tiedot kerättiin, sekä b) **kuvailevaa tietoa**

siitä, miten aineistoa ja dataa tulkitaan oikein metadatastandardien, sanastojen ja esimerkiksi readme-tiedostojen avulla.

#### Vinkkejä hyvistä käytännöistä

- Kuva kaikki tarjoamasi dokumentaatiotyypit (esim. README-tiedostot ja metatiedot), jotka auttavat jatkokäyttäjiä löytämään, ymmärtämään ja uudelleen käyttämään aineistoa. Arkistot edellyttävät usein tietyn metatietostandardin käyttöä. Tarkista, onko olemassa tieteenalakohtaista metadatumallia tai -standardia, jonka voisit ottaa käyttöön.
- Pohdi, miten tiedot järjestetään projektin aikana. Kuvaile esimerkiksi tiedostojen nimeämiskäytäntöjä, versionhallintaa ja kansiorakennetta.
- Käytä välineitä ja laitteita, jotka tuottavat standardoitua metadataa.
- Tunnista ne vähimmäistason kuvailutiedot, jotka tulee tallentaa, jotta muut tutkijat voivat löytää, tarkastella, tulkita ja käyttää aineistoasi sekä viitata siihen. Lisätietoja aineiston dokumentaation vähimmäisvaatimuksista on esimerkiksi täällä: (<https://www.fairdata.fi/qvain/qvain-light-kayttoopas/>).

## 4. Tallentaminen ja varmuuskopiointi tutkimushankkeen aikana

### 4.1 Minne aineistosi tallennetaan, ja miten se varmuuskopioidaan?

Kuvaa, minne aineisto tallennetaan ja varmuuskopioidaan tutkimushankkeen aikana. Pohdi, kuka vastaa tietojen varmuuskopioinnista ja palauttamisesta. Jos mukana on useita tutkijoita, tee suunnitelma yhteistyökumppaniesi 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.

Kuvaa kohdassa 5 yksityiskohtaisesti menetelmät, joita käytetään aineiston säilyttämiseen ja jakamiseen tutkimushankkeen päätyttyä.

#### Vinkkejä hyvistä käytännöistä

- On suositeltavaa käyttää oman organisaatiosi tai muun luotettavan IT-tarjoajan (kuten CSC) tarjoamaa ja ylläpitämää tietoturvallista tallennusratkaisua.
- ÄLÄ KÄYTÄ ulkoisia kovalevyjä pääasiallisena säilytysratkaisuna.
- Noudata organisaatiosi aineiston turvallisen tallentamisen vaatimuksia.

### 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.

#### Vinkejä hyvistä käytännöistä

- Aineistoon pääsyä tulee aina valvoa luottamuksellisuustason 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?

Kuvaa, miten saatat aineistosi löydettäväksi ja saataville jatkokäyttöä varten. Jos aineistoa tai osaa siitä ei voi julkaista, perustele, miksi julkaiset ainoastaan metadatan.

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.

#### Vinkejä 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/) (<https://www.re3data.org/>).
- Suosi sellaisten arkistojen tai julkaisijoiden käyttämistä, jotka tarjoavat pysyvän tunnisteiden (PID). Näin varmistat, että aineisto on käytettävissä pysyvän linkin (esim. DOI, URN) kautta.
- 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.

**VÄLTÄ PÄÄLLEKKÄISYYTTÄ JULKAISUSUUNNITELMAN KANSSA!** Tutkimusartikkelin julkaiseminen ei ole sama asia kuin aineiston julkaiseminen. Datajulkaisut (*data journal*) ovat tutkimusaineistojen julkaisuun erikoistuneita foorumeita.

#### 5.2 Missä pitkällä aikavälillä arvokas aineisto säilytetään ja kuinka pitkään?

Kuvaa lyhyesti, mitkä osat aineistostasi aiot säilyttää, missä ja kuinka pitkään. Pitkäaikaissäilytys tarkoittaa, että aineisto ja data säilytetään niin pitkään kuin on tarpeellista - jopa useiden vuosikymmenten tai -satojen ajaksi.

Voit luokitella aineistosi ennakoidun 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

Sinun on päätettävä, mitkä aineistot ja data säilytetään ja mitkä hävitetään. Ainutlaatuisella tai vaikeasti toistettavalla datalla voi olla pitkäaikaista arvoa ja se on syytä säilyttää. Digitaaliseen säilyttämiseen tulisi käyttää erityisiä pitkäaikaiseen säilyttämiseen erikoistuneita arkistoja ja data repositorioita.

#### Vinkkejä hyvistä käytännöistä

- Aineistojen ja datan säilyttämistä koskevat päätökset tulisi tehdä jo aineistohallinnan suunnitteluvaiheessa, ja niiden tulisi ottaa huomioon mm. organisaatioiden ohjeet ja vaatimukset.
- Käytä arkistoja, jotka ovat sitoutuneet aineistojen ja datan pitkäaikaiseen kuratointiin. Esimerkiksi [Fairdata PAS-palvelu](#) on erikoistunut tutkimusaineistoihin, joilla on merkittävää arvoa organisaatiolle tai kansallisella tasolla sekä tällä hetkellä että erityisesti tulevaisuudessa. Lisätietoa PAS-palvelusta - ota yhteyttä kotiorganisaatiosi datatukeen.

## 6. Aineistohallintaa koskevat vastuut ja resurssit

### 6.1 Kuka (esim. tehtävä ja laitos) vastaa aineistohallinnasta?

Kerro tässä lyhyesti, miten edellisissä vastauksissa kuvatut tehtävät ja vastuut on jaettu. Pohdi myös, kuka vastaa aineistosta, jota hankkeesi tuottaa vielä päätyttyään.

#### Vinkkejä hyvistä käytännöistä

- Hahmottele aineistohallintaan-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 aineistohallintavastuut.
- Määrittele, kuka vastaa aineistohallintasuunnitelman toteuttamisesta ja varmistaa, että suunnitelma tarkastetaan ja tarvittaessa korjataan.
- Harkitse aineistohallintasuunnitelman säännöllistä päivittämistä.

## **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. Tarvittavia resursseja ovat esimerkiksi säilytyskustannukset, laitteet, työaika sekä mahdolliset data-arkiston tallennuskustannukset.

### **Vinkejä hyvistä käytännöistä**

- Pohdi, mitä täydentäviä tietoteknisiä tiloja ja resursseja tarvitaan ja mitkä niiden kustannukset ovat.
- Harkitse tarkkaan, mitä resursseja tarvitaan aineiston jakamiseen, säilyttämiseen ja kuratointiin.
- Erittele aineistonhallintakustannukset budjetissa rahoittajan vaatimusten mukaisesti.