

Tiedeinstituuttien avoin tutkimusdata

Digitaalisen aineistonhallinnan käsikirja

Versio 1.1

Harri Kiiskinen, Laura Nissin & Manna Satama

2020-12

Digitaalisen aineistonhallinnan käsikirja, versio 1.1

2020-12

Harri Kiiskinen, <https://orcid.org/0000-0003-4187-5551>

Laura Nissin, <https://orcid.org/0000-0001-8050-3916>

Manna Satama, <https://orcid.org/0000-0003-3775-9363>

Tiedeinstituuttien avoin tutkimusdata -hanke

Institutum Romanum Finlandiae sr

Suomen Ateenan-instituutin säätiö sr

Suomen Lähi-idän instituutin säätiö sr

Suomen Japanin-instituutin säätiö sr

DOI: <https://doi.org/10.5281/zenodo.4327541>

Asiasanat (YSA): avoin tieto; metadata; tutkimusaineisto; humanistiset tieteet; kuvailu; projektit; pysyvät tunnisteet; tiedeinstituutit; tiedontuottajat; tieteellinen julkaisutoiminta; tutkimusprojektit; tutkimustyö

CC BY 4.0

© 2020 Tiedeinstituuttien avoin tutkimusdata -hanke

Tämä asiakirja on tehty osana opetus- ja kulttuuriministeriön rahoittamaa Tiedeinstituuttien avoin tutkimusdata -hanketta. Hankkeen toteuttajina toimivat Suomen ulkomaisten tiedeinstituuttien taustayhteisöt. Hanketta hallinnoi *Institutum Romanum Finlandiae sr.*

Asiakirja on hankkeen tutkijoiden ehdotus instituuttien työntekijöille ja tutkijoille suunnatusta ohjeistuksesta. Se koskee digitaalisten tutkimusaineistojen hallintaa, tuottamista, omistajuutta, säilyttämistä ja julkaisua. Asiakirjassa esitetyt näkemykset perustuvat hankkeen tutkijoiden selvityksiin. Sitä on käsitelty hankkeen ohjausryhmässä, joka on antanut sisältöön liittyviä ohjeita, kommentteja ja palautetta.

Asiakirjan ensimmäinen versio 1.0 julkaistiin lokakuussa 2020. Siihen oli kerätty palautetta erityisesti hankkeessa mukana olevien instituuttien henkilöstöltä ja instituutteja ylläpitävien säätöiden hallituksilta kesällä ja syksyllä 2020. Käsillä oleva versio 1.1. on edellisestä versiosta palautteiden perusteella päivitetty asiakirja. Keskeinen muutos on edellistä versiota laajempi johdanto-osa, jossa asiakirjan tarkoitusta ja keskeisiä käsitteitä avataan enemmän.

Asiakirja on tarkoitettu tiedeinstituuttien käyttöön niiden sopivaksi katsomalla tavalla.

SISÄLLYS

| | |
|---------------------------------------------------------------------------|----|
| Tiivistelmä | ii |
| 1. Johdanto | 1 |
| Käsikirjan kohderyhmä ja tavoitteet | 1 |
| Avoin tiede ja tutkimus tutkijan näkökulmasta | 2 |
| FAIR-periaatteet | 4 |
| 2. Perusasiaa aineiston hallintaan ja avoimeen tutkimukseen | 5 |
| Mitä on aineisto? Mitä on data? | 5 |
| Mitä on metatieto tai metadata? | 5 |
| Seitsemän askelta aineistohallintaan FAIR-periaatteet huomioiden | 6 |
| 3. Miten toimitaan, kun tullaan instituuttiin tekemään tutkimusta | 7 |
| 1) Aineistohallintasuunnitelma | 7 |
| 2) Data-arkiston valinta | 8 |
| 3) Järjestely ja tiedostomuodot | 8 |
| 4) Dokumentointi | 9 |
| 5) Kuvailu: metatiedot, sanastot, alustat | 9 |
| Dublin Core (DC) | 9 |
| Data Documentation Initiative (DDI) | 10 |
| Gettyn sanastot | 10 |
| Finto eli suomalaiset asiasanastot | 10 |
| CIDOC-CRM eli Conceptual Reference Model | 10 |
| Wikidata – Avoin tietoaarkisto | 10 |
| Kuvailun työkaluja | 10 |
| 6) Tutkimuksenaikainen säilytys | 11 |
| Aineistohallintajärjestelmiä | 11 |
| 7) Valmis tutkimusaineisto | 12 |
| Data-arkisto eli repositorio | 12 |
| Yleinen, yksilöimätön käyttö lupa, esim. Creative Commons -lisenssi | 12 |
| Yksilöity käyttö lupa | 13 |
| Pysyvä tunniste eli PID | 13 |
| A. LIITE: Instituuttien aineistohallintasuunnitelman malli | 14 |
| B. LIITE: README-tiedosto | 16 |
| C. LIITE: Dublin Core (erillinen tiedosto) | |

1. JOHDANTO

Käsikirjan tarkoitus on ohjeistaa instituuteissa työskentelevien tutkijoiden tutkimusaineiston hallinnan käytäntöä mahdollisimman konkreettisesti. Käsikirjassa on huomioitu aiheen kansainvälinen ja kansallinen keskustelu ja tämänhetkinen toiminta. Käsikirjassa esitettyjen ratkaisujen avulla on mahdollista pyrkiä avoimen datan FAIR-periaatteiden mukaiseen toimintaan tutkimusaineistojen hallinnassa ja avaamisessa.

Käsikirja auttaa tutkijaa toteuttamaan

- aineistohallintasuunnitelman (ja muistuttaa myös sen päivittämisen tärkeydestä)
- tutkimuksenaikaisen aineistohallinnan perustan
- riittävän avoimen aineiston FAIR-tavoitteiden mukaisesti eli niin avointa kuin mahdollista, niin suljettua kuin pakollista
- aineiston sellaiseen kuntoon, että sen voi toimittaa julkaisu- tai data-arkistoon (repositorioon) mahdollisimman vaivattomasti.

Johdantoluvussa käydään läpi avoimen tieteen perusasioita ja FAIR-periaatteita erityisesti tutkijoiden näkökulmasta.

KÄSIKIRJAN KOHDERYHMÄ JA TAVOITTEET

Käsikirjan ensisijainen kohderyhmä ovat instituuttien tutkijat ja tutkimushankkeiden johtajat. Se on hyödyksi myös tutkimuksen hallinnon parissa työskentelevälle henkilöstölle ja jopa muusta hallinnosta vastaaville ja instituutteja ylläpitävien säätiöiden hallituksen jäsenille. Instituutin tutkijalla tarkoitetaan tässä yhteydessä erityisesti instituutin palkkalistoilla tai apurahalla olevaa henkilöä tai vastuuhenkilöä tutkimusprojektissa, jonka rahoitukseen ja/tai toteuttamiseen instituutti osallistuu.

Käsikirja on tarkoitettu myös instituuttien piiriä laajemmalle tutkija- ja opiskelijajoukolle. Sitä voi vapaasti käyttää ja muokata eri tutkimuslaitosten tarpeisiin vaikkapa opetuksessa ja tutkimuksen tuessa. Yksittäisen tutkijan kohdalla käsikirjan hyöty näyttäytyy siinä, että se ohjaa itsenäiseen aineistohallinnan toteuttamiseen tilanteessa, jossa tutkijalla ei ole taustaorganisaation tarjoamaa laaja-alaista tutkimuksen tukea. Käsikirja on laadittu juuri sellaisen tutkimusorganisaation piirissä, jossa tutkimusinfrastruktuuri on vaatimaton ja jonka mahdollisuudet hyödyntää kansallisia palveluita nykytilanteessa ovat rajoitetut organisaatiosta riippumattomista syistä. Yksityisinä toimijoina tiedeinstituutit eivät esimerkiksi voi käyttää pääosin valtion rahoittaman ja suomalaisten yliopistojen osaomistuksessa olevan CSC:n eli Tieteen tietotekniikan keskus Oy:n palveluita lainkaan.¹

Käsikirjan tavoitteena on esitellä konkreettisia, pieniä askeleita avoimuuden toteuttamiseen. Se antaa eväitä koko tutkimusprosessin aineistohallinnan järjestämiseen. Se tarjoaa ohjeistuksen lisäksi mallit aineistohallintasuunnitelmaan ja README-tiedostoon sekä esittelee monen tyyppisessä käytössä olevan metatietomallin (Dublin Core).

¹ CSC:n ja instituuttien välistä tilannetta käsitellään Tiedeinstituuttien avoin tutkimusdata -hankkeen loppuraportissa: Kiiskinen, Nissin & Satama 2020, *Loppuraportti – Tiedeinstituuttien avoin tutkimusdata -hanke 2018–2020* (<https://doi.org/10.5281/zenodo.4323115>).

AVOIN TIEDE JA TUTKIMUS TUTKIJAN NÄKÖKULMASTA

Käsillä olevan asiakirjan perusta ja keskeinen teema on sitoutuminen avoimen tieteen ja tutkimuksen periaatteisiin ja käytäntöihin. Avoimen tieteen periaatteilla viitataan yleisesti toimintamalleihin, joilla edistetään tieteen avoimuutta. Tieteeseen kuuluu erottamattomasti avoimuus, mutta käytännössä tämä ei aina toteudu vaikkapa silloin, kun tutkija julkaisee tulokset maksumuurin takana tai vaikeasti löydettävissä olevassa julkaisussa tai kun tutkimustulokseen johtaneita aineistoja tai menetelmiä ei ole dokumentoitu siten, että ne ovat avoimesti saatavilla. Avoimen tieteen ja tutkimuksen keskustelu liittyy siten tutkimuksen eri osatekijöiden avoimen saatavuuden edistämiseen ja toteuttamiseen aina aineistoista ja menetelmistä julkaisuihin ja oppimateriaaleihin.²

Keskustelu liittyy oleellisesti myös tieteen eettisiin periaatteisiin. Tutkijalla on oikeus ja vastuu suunnitella ja toteuttaa tutkimusaineistonsa hallinta tieteen etiikka huomioiden. Eettisistä kysymyksistä löytyy lisätietoa Tutkimuseettisen neuvottelukunnan sivulta ([TENK](#)).

Sellaisenaan avoin tiede ja tutkimus saattaa tuntua itsestään selvältä ja jopa yksinkertaiselta. Kun tavoitetta puretaan käytännön elämään, esiin nousee ongelmallisia ja vaikeita seikkoja, jotka liittyvät kaikkiin tieteellisen toiminnan tekijöihin ja eri näkökulmiin. Niiden ymmärtäminen on edellytys sille, että totuttuja toimintatapoja voidaan muuttaa.

- Tutkijan näkökulmasta ongelmiksi voivat nousta osaaminen, motivaatio, ajan ja rahan puute, ymmärrys, ”ylhäältä sanelu”, työsuhteet, aineistojen omistajuus ja tekijänoikeudet sekä meritoituminen.
- Tutkimusyhteisössä jaetaan yksittäisen tutkijan kysymyksiä ja ongelmia. Tarvitaan siis koko yhteisön tuki ja osallistuminen toimintatapojen muutokseen. On hyväksyttävä eräänlainen paradigman muutos, jossa tutkimuksen näkyvyys ei painotu ainoastaan ns. loppujulkaisuun vaan tutkimuksen koko prosessiin.
- Rahoittajan rooli on merkittävä. Rahoitusmekanismit ohjaavat käytäntöjä. Jos edellytetään datan hallintaa ja avaamista, sitä yleensä yritetään tehdä, jotta saataisiin rahoitusta. Rahoittajan pitää investoida lisää yksittäisiin tutkimuksiin, jotta datan hallinta mahdollistetaan käytännössä. Tärkeää: Raha ei saa valua tutkimuksesta hallintoon! Tutkimusrahoitusta haettaessa on muistettava ja uskallettava hakea riittävästi rahaa tutkimusaineiston hallintaan ja rahoittajan tulee huomioida tämä eräänlaisena lisärahoituksena.
- Tutkimusorganisaatiolta edellytetään johdon sitoutumista. On sitouduttava avoimen tieteen periaatteiden ajamiseen ja siihen on myös budjetoitava. Tämä edellyttää luottamusta tutkijoiden ja organisaation johdon välillä.

² Avoimen tieteen ja tutkimuksen keskustelua, kansallisia ja kansainvälisiä linjauksia ja ohjeita löytyy kootusti TSV:n yhteydessä toimivan Avoimen tieteen kansallisen koordinaation *Avoim tiede* -sivustolta: <https://avointiede.fi/fi>.

Tutkijan näkökulmasta on perusteltua sitoutua avoimen tieteen ja tutkimuksen periaatteiden edistämiseen tutkimusaineistojen osalta muistakin kuin tieteen yleisistä, eettisistä näkökulmista.

- **Meritoituminen:** Tutkimusaineistojen julkaiseminen niin avoimesti kuin lainsäädännöllisesti ja eettisesti mahdollista on meriitti työuralla yhä enenevässä määrin. Meritoitumiseen vaikuttava toimintakulttuuri muuttuu, vaikka hitaasti. Pysyvien tunnisteiden käyttö tutkimusaineistojenkin kohdalla mahdollistaa mm. sen, että muut tutkijat voivat niihin viitata, mikä puolestaan lisää tutkijan ja hänen tutkimuksensa näkyvyyttä.
- **Rahoitus:** Tutkimusrahoittajat edellyttävät yhä useammin sekä aineistojen että tutkimustulosten avointa jakamista. Avoimen tieteen periaatteiden edistäminen uskottavasti omassa tutkimuksessa on etu rahoitusta haettaessa.
- **Tutkimus itsessään:** Tutkimuksenaikainen aineistonhallinta jäntevoityy, kun seuraa avoimen tieteen mahdollistavia käytännön toimenpiteitä ja ohjeita.

FAIR-PERIAATTEET

FAIR on akronyymi sanoista Findable, Accessible, Interoperability ja Reusability. FAIR-periaatteet viittaavat siihen, että aineisto on löydettävissä (Findable), siihen päästään käsiksi (Accessible), se toimii yhdessä eri järjestelmien ja hakukoneiden kanssa (Interoperability) ja sitä voidaan käyttää uudelleen (Reusability). Vaikka FAIR-periaatteita pidetään toisinaan saavuttamattomina toiveina, niihin kannattaa silti suhtautua tavoitteellisesti. Ne ohjaavat tutkimusaineiston käsittelyä ja hallintaa suunnittelusta arkiseen toteuttamiseen sekä tutkimuksen hallinnossa että itse tutkimuksessa.

Fairdata.fi -palvelu tarjoaa päivitettyä ja luotettavaa tietoa datan hallinnasta FAIR-periaatteet huomoiden. Palvelun järjestää opetus- ja kulttuuriministeriö ja toteuttaa CSC – Tieteen tietotekniikan keskus Oy (<https://www.fairdata.fi/>). Tässä kohdin on otettu esiin muutamia perusasioita, joiden tietoisella huomiolla on jo otettu tärkeitä askelia FAIR-periaatteiden ja avoimen tieteen käytännön toteuttamisessa. Alla oleva taulukko pohjautuu Fairdata.fi -sivustolta löytyvään tietoon.

| FAIR-periaatteet | Mitä |
|----------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Findable LÖYDETTÄVYYS | Aineisto saa pysyvän, ainutkertaisen tunnisteen (PID, persistent identifier) (esim. doi) ja tunniste mainitaan metatiedoissa. Aineistolle annetaan kattavat metatiedot . Metatieto on <ul style="list-style-type: none"> - kuvailevaa (aineiston sisältö) - hallinnollista (esim. käyttöehdot, käytön edellyttämät tekniset vaatimukset) - rakenteellista (esim. tiedostojen suhde toisiinsa) Aineisto kuvaillaan ja rekisteröidään hakupalveluun (esim. tieteenalakohtaiset tai kansalliset palvelut). |
| Accessible SAAVUTETTAVUUS | Aineisto tai sen metatiedot ovat noudettavissa standardisoidun yhteyskäytännön kautta. Tavoitteena on niin avoin aineiston saatavuus kuin mahdollista ja niin suljettuja kuin pakollista. Yhteyskäytäntö on avoin, maksuton ja yleisesti käytettävissä . Metatiedot ovat saavutettavia , vaikka itse aineisto ei (enää) olisikaan. |
| Interoperability YHTEENTOIMIVUUS | Aineisto on sekä ihmis- että koneluettavaa . Aineistossa ja metatiedoissa käytetään sanastoja ja ontologioita , jotka ovat koneluettavia, avoimia ja yleisesti käytössä. Mainitaan aineiston suhde muihin aineistoihin, esim. eri versiot tai viittaukset. |
| Reusability UUDELLEEN KÄYTETTÄVYYS | Aineistolle on annettu kattavat metatiedot . Aineiston käytön ehdot (lisenssit, sopimukset) on mainittu selkeästi. Aineiston alkuperä ja syntyprosessi on kerrottu selkeästi. |

FAIR-periaatteiden mukaista toimintaa on tutkimuksen jokaisessa vaiheessa. Suuri vastuu on tutkijalla, joka on aineistonsa paras asiantuntija ja sitä kautta vastaa pitkälti esim. aineiston kuvailusta, käyttöehtojen kirjaamisesta, sopivan julkaisuarkiston käyttämisestä. Tutkijat ovat avainasemassa, kun puhutaan toimintakulttuurin muutoksesta ja suhtautumisesta tutkimusaineistojen julkaisemiseen viittaamalla omaan ja muiden tutkimusaineistoihin sekä hyödyntämällä jo julkaistuja tutkimusaineistoja. Tutkijan työnantajan tehtävä on tukea tätä muutosta tutkimuksen eri vaiheissa alkaen rekrytoinnista.

2. PERUSASIAA AINEISTON HALLINTAAN JA AVOIMEEN TUTKIMUKSEEN

MITÄ ON AINEISTO? MITÄ ON DATA?

Aineisto on tutkimuksen sisällöllinen perusta. Aineisto voi olla analogisessa tai digitaalisessa muodossa. Usein digitaalista aineistoa kutustaan dataksi.

Aineistoa ovat esimerkiksi

- mittaukset
- paikkatiedot
- kuvat
- äänitteet
- tekstit
- luettelot
- muistiinpanot
- sähköpostit
- tutkijan luoma uusi aineisto toisen aineiston pohjalta

MITÄ ON METATIETO TAI METADATA?

Metatieto kuvailee aineiston tai datan sen koko elinkaaren ajan. Digitaalista metatietoa kutsutaan usein metadataksi.

Metatieto huomioi aineiston

- kontekstin
- sisällön
- rakenteen
- hallinnan
- käsittelyn.

Käytännössä tällaisia tietoja ovat esimerkiksi aineiston nimi, tuottamisajankohta, tuottaja, muoto, aihe ja käyttöoikeus.³ On hyvä muistaa, että vaikka tutkimusaineistoa ei jostain syystä voi avata, metatieto voidaan avata.

Metatiedolle suositellaan standardeja, joita on lukuisia eri tieteenaloja ja käyttötarkoituksia varten. Standardien käyttö edistää löydettävyyttä, saavutettavuutta, yhteentoimivuutta ja uudelleenkäyttöä eli johdantoluvussa kuvailtujen [FAIR](#)-periaatteiden toteumista.

³ Esim. Turun yliopiston ohjeistus, <<https://utuguides.fi/tutkimusdata/avaaminen#s-lg-box-14879918>>; Tietoarkiston *Aineistonhallinnan käsikirja*, <<https://www.fsd.tuni.fi/fi/palvelut/aineistonhallinta/aineiston-kuvailu-ja-metadata/>>.

SEITSEMÄN ASKELTA AINEISTONHALLINTAAN FAIR-PERIAATTEET HUOMIOIDEN

Nämä seitsemän kohtaa nostavat esiin keskeiset asiat, joista huolehtimalla avoimen tieteen tavoitteet konkretisoituvat ja ovat hallittavissa omassa tutkimuksessa suunnitelmallisesti. Tässä yhteydessä kukin kohta tai askel kuvaillaan tai perustellaan lyhyesti sillä ajatuksella, että listamaiseen esitystapaan on helppo palata tutkimuksen aikana. Käsikirjan luvussa kolme kuhunkin kohtaan paneudutaan tarkemmin ja annetaan esimerkkejä ja linkkejä työkaluihin ja palveluihin.

1) [Suunnittele](#): Aineistonhallintasuunnitelma eli DMP

Suunnittele aineiston hallinta, palaa siihen ja päivitä suunnitelma. Käytä aikaa aineistonhallintasuunnitelman tekemiseen ja seuraamiseen aivan kuten käytät aikaa tutkimussuunnitelman laatimiseen, tutkimuksen raportointiin tutkimuksen aikana ja tulosten julkaisemiseen.

2) [Valitse](#) data-arkisto eli repositorio.

Valitse data-arkisto, johon suunnittelet tallentavasi aineiston. Tutustu sen ohjeisiin ja päivitä tarvittaessa aineistonhallintasuunnitelma valitun data-arkiston ohjeiden mukaisesti.

3) [Järjestele aineisto ja huomioi tiedostomuodot](#).

Pidä aineisto aineistonhallintasuunnitelmassa kuvatussa järjestyksessä ja suosi yleisiä tiedostomuotoja jatkokäytön turvaamiseksi.

4) [Dokumentoi: README-tiedosto](#).

Tee README-tiedosto, joka kulkee aineiston mukana.

5) [Kuvaile: metadata](#).

Aineiston metatiedot turvaavat sen ymmärrettävyyden.

6) [Säilytä: varmuuskopiot ja versiot](#).

Tutkimuksenaikaisessa säilytyksessä on muistettava varmuuskopiointi, versiointi ja tarpeen mukaan jakaminen.

7) [Päätä valmiin tutkimusaineiston kohtalo](#).

Tutkimukseen kuuluu loppujulkaisun rinnalle aineistoon kohdistuvat lopputyöt eli mitä säilytetään, kuinka kauan ja missä sekä millä tavoin sitä jaetaan. Keskeisiä käsitteitä ovat data-arkistot, lisenssit, käyttöluvut ja pysyvät tunnisteet.

3. MITEN TOIMITAAN, KUN TULLAAN INSTITUUTTIIN TEKEMÄÄN TUTKIMUSTA

Aineistohallinnassa ja erityisesti teknisiä ratkaisuja valittaessa on muistettava huolehtia tietosuojasta. Tietosuojavaltuutetun toimiston sivustolla on selkeä esitys henkilötiedoista ja tietosuojasta ([Tietosuoja](#)).

1) AINEISTONHALLINTASUUNNITELMA

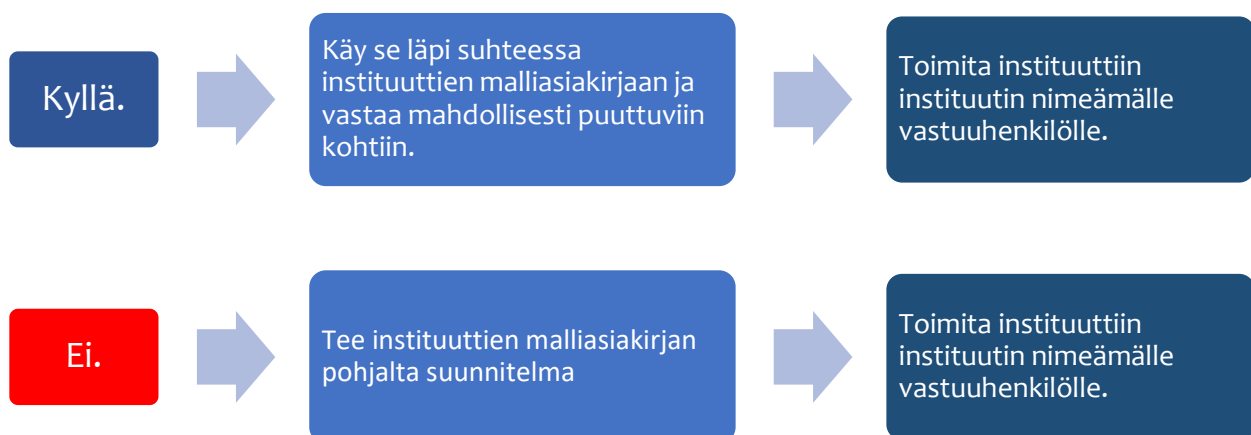
Aineistohallintasuunnitelmaan viitataan usein lyhenteellä DMP eli Data Management Plan.

Käsikirjan liitteenä on instituuttien aineistohallintasuunnitelman malliasiakirja ([Liite A](#)). Tarjolla on myös runsaasti muita ohjeita ja työkaluja aineistohallintasuunnitelman tekemiseen, esimerkiksi:

- [DMPTuuli](#) on suomalainen työkalu aineistohallintasuunnitelman laatimiseen. Sen käyttö ei edellytä kuulumista esimerkiksi johonkin yliopistoon tai julkishallinnon tutkimusorganisaatioon. Sivustolta löytyy myös eri organisaatioiden asiakirjapohjia ja julkaistuja aineistohallintasuunnitelmia.
- [easyDMP](#) on EU:n EUDAT-palvelun tuottama työkalu aineistohallintasuunnitelman laatimiseen. EUDAT-tutkimusinfrastruktuuria käsitellään lyhyesti myöhemmin kohdassa 6 tutkimuksenaikaisen aineiston säilytyksen yhteydessä.

Ensimmäinen kysymys, kun tullaan instituuttiin tekemään tutkimusta:

Onko jo olemassa aineistohallintasuunnitelma?



2) DATA-ARKISTON VALINTA

Data-arkiston eli repositorion alustava valinta on hyvä tehdä jo varhain, jos vain mahdollista. Usein arkistoilla on ohjeet, joiden mukaan aineisto pitää järjestellä ja muotoilla. Siksi arkiston valinta jo tutkimuksen alkuvaiheissa voi helpottaa merkittävästi aineistonhallinnan käytännön ratkaisuisissa, koska ohjeita noudattaen aineisto on varmasti arkistokelpoinen. Data-arkistojen sivuilta voi löytyä muutoinkin hyviä ohjeita aineiston järjestämiseen ja käsittelyyn.

Suomessa esim. Tietoarkisto ottaa vastaan humanistiseen tutkimukseen liittyviä aineistoja ja antaa ohjeet sen järjestelyyn. EU-tason peruspalvelu on Zenodo.org, joka ottaa vastaan kaikkea tutkimusaineistoa ilmaiseksi eikä aseta juurikaan vaatimuksia aineiston järjestämiselle. Alakohtaisia data-arkistoja voi etsiä mm. [Registry of Reserach Data Repositories -työkalun](#) avulla. Monet arkistoista ovat kansallisia tai maksullisia.

Tarjolla olevien palveluiden kenttä on nopeasti muuttuva, joten pysyviä suosituksia on vaikea antaa. Esimerkiksi CSC eli Tieteen tietotekniikan keskus Oy kehittää uusia [kansallisia ratkaisuja](#), jotka saattavat olla käyttökelpoisia ja suositeltavia joidenkin vuosien kuluessa.

3) JÄRJESTELY JA TIEDOSTOMUODOT

Kiinnitä jo heti alkuun huomiota aineiston keskinäiseen järjestykseen ja perustele se. Dokumentoi järjestys myös README-tiedostoon (ks. seuraava kohta 4).

Viimeistään siinä vaiheessa, kun aineisto siirretään data-arkistoon, kiinnitä huomiota tiedostomuotoihin. Tämä varmistaa osaltaan aineistosi käytettävyyden.

Ohjenuora:

Suosi avoimen lähdekoodin tiedostomuotoja. Vältä omisteisia ohjelmia.

Alla olevassa taulukossa annetaan esimerkkejä suositeltavista ja vältettävistä tiedostomuodoista.

| Tyyppi | Suosi | Vältä |
|---------------------------------------|---------------------------------------------------|-----------------|
| Teksti | plaintext/markdown, xml/html, odt, pdf/a | word |
| Taulukko | CSV, TSV, SPSS portable | excel |
| Media | Container: mp4, ogg Codec: Theora, Dirac, FLAC | Quicktime, HZ64 |
| Kuva | DNG, TIFF, JPG 2000, PNG | GIF, JPG |
| Rakenteinen data (structured data) | xml, rdf, json, yaml | RDBMS, Access |

4) DOKUMENTOINTI

Aineiston mukana kulkeva aputiedosto turvaa tutkimusaineiston ymmärrettävyyden. Siinä kerrotaan aineiston keskeiset seikat, kuten aineiston nimi, tekijä, käyttöoikeudet, kuvailustandardit jne.

Tällainen aputiedosto kannattaa nimetä README-tiedostoksi, jotta sen erottaa aineiston muiden tiedostojen joukosta. Tallenna tiedosto tekstitiedostona, esimerkiksi txt-muotoisena eikä esimerkiksi word-tiedostona. Varmista myös, että README-tiedoston kulkee aineiston mukana.

README-tiedoston malli on *Käsikirjan* [liitteessä B](#).

5) KUVAILU: METATIEDOT, SANASTOT, ALUSTAT

Tärkeintä on noudattaa tutkimushankkeen kannalta toimivinta kuvailun tapaa. Tarkoitus ei ole pakottaa tutkimusta valmiisiin standardeihin, jos ne eivät ole tarkoituksenmukaisia. Kuvailun selittäminen on kuitenkin ehdottoman tärkeää, käytetään standardia tai ei. Tutkimushankkeessa kaikkien on ymmärrettävä, miten aineistoa kuvaillaan termien sisällöstä alkaen.

Ohjenuora:

Suosi standardoituja tapoja ja sanastoja ilmaisemaan kentissä kerrottuja asioita.

Esimerkiksi tekijän nimi on tyyliä ”Sukunimi, Etunimi” ja päivämäärä ”vuosi-kuukausi-päivä”.

Jos käytössä on tietty metatietostandardi, muista mainita README-tiedostossa, mikä standardi on käytössä, tai jos standardia ei käytetä, missä kuvailutietojen logiikka selitetään.

Standardit määrittelevät kolmea eri metatiedon tasoa: käytettävissä olevia tietokenttiä (”schemes”), kentissä käytettäviä sanastoja (”taxonomies and vocabularies”) ja eri tietokenttien ja aineistoyksiköiden välisiä suhteita (”ontologies”).

Alla on kerrottu joitain metatietostandardeja sekä kuvailuun liittyviä työkaluja ja alustoja.

DUBLIN CORE (DC)

Matalan kynnyksen skeema on [Dublin Core](#) (DC), jota käytetään laajasti ja jota on sillattu eli mäpätty muihin alakohtaisiin standardeihin ja malleihin. DC:n plussana on sen monikäyttöisyys: Se taipuu eri tieteenaloihin ja aineistoihin, sillä sen ns. peruskentät ovat valtavan yleisluonteisia ja helposti ymmärrettäviä ja niistä voi valita tarpeelliset kentät. Aineiston kuvailun voi aloittaa jo tutkimuksen alussa ja tarkentaa tutkimuksen aikana. DC:n kentät ohjaavat myös vastaamaan moniin aineistonhallintasuunnitelmassa esitettyihin kysymyksiin.

DC-standardia voikin pitää eräänlaisena kattotason jäsentelyn apuna tai aineiston julkaisemisen yhteydessä aineiston kokonaisuuden kuvailun työkaluna. *Käsikirjan* liitteessä C on annettu Dublin Coren kentät (eng Elements) ja kenttätarkenteet suomeksi ja englanniksi sekä viittaukset käytettäviin sanastoym. standardeihin ja suosituksiin. Liite on erillisenä taulukkotiedostona. Dublin Core-metatietoja varten on myös olemassa [Dublin Core -generaattori](#).

DATA DOCUMENTATION INITIATIVE (DDI)

[DDI](#)-metatietostandardia käytetään erityisesti yhteiskunta- ja käyttäytymistieteissä. Se soveltuu hyvin kyselytutkimuksiin. Standardista on kaksi versiota, joista DDI-Lifecycle on tarkoitettu nimensä mukaisesti aineiston koko elinkaarelle ja DDI-Codebook jossain määrin kevyempään käyttöön. Esimerkiksi Tietoarkisto käyttää DDI Codebook -versiota. DDI:n versiot ovat rakenteista, koneluettavaa ja xml-pohjaista soveltuen siten pitkäaikaisäilytykseen.

DDI-standardissa on viisi pääkenttää ja lukuisa määrä näiden alla olevia lisäkenttiä. Se on sillattu moniin muihin standardeihin, ja esimerkiksi valtaosa Dublin Coren kentistä saa vastineen DDI:n kentistä, kuten yksinkertainen [DC-DDI -taulukko](#) osoittaa. Kenttien ja niiden alakategorioiden kuvakset löytyvät [DDI:n kotisivulta](#).

GETTYN SANASTOT

Getty Research Institute ylläpitää taiteeseen ja kulttuuriperintöön liittyviä monikielisiä sanastoja, joiden kattamat terminologiat soveltuvat hyvin instituuttien piirissä tehtävän tutkimuksen aineistojen kuvailuun (<https://www.getty.edu/research/tools/vocabularies/>).

FINTO ELI SUOMALAISET ASIASANASTOT

”Finto on suomalainen sanasto- ja ontologiapalvelu, joka mahdollistaa sanastojen julkaisun ja selailun. Palvelu tarjoaa myös rajapinnat sanastojen ja ontologioiden hyödyntämiseen muissa ohjelmistoissa” (<http://finto.fi/fi/>).

CIDOC-CRM ELI CONCEPTUAL REFERENCE MODEL

[CIDOC-CRM](#) (Conceptual Reference Model) on erityisesti kulttuuriperintöaineiston kuvailuun luotu ontologia. Sen avulla kuvataan tallennettavan tiedon sisäisiä suhteita. Laajennuksineen ontologia soveltuu esimerkiksi arkeologisen kaivausprojektin kaikkien vaiheiden kuvailuun.

CIDOC-CRM:n mahdollistaa konkreettisen tai abstraktin kohteen koko elinkaaren kuvailun. Sen soveltaminen aineistojen kuvailuun ei ole aivan yksinkertaista, mutta potentiaalisesti se on hyvin hyödyllistä aineiston pitkäaikaiskäytön kannalta.

WIKIDATA – AVOIN TIETOARKISTO

[Wikidata](#) on avoin tietoaarkisto, jonka tarkoitus on toimia sekä ihmisen että koneen luettavissa ja muokattavissa olevan tiedon avoimena tallennusalueena. Wikidata ei suoraan sovellu tutkimusaineiston kuvailuun, mutta siitä voi olla hyötyä tutkimusten tulosten julkaisussa tai tutkimusdatan kokonaisuuden kontekstualisoinnissa.

KUVAAILUN TYÖKALUJA

Kenttätöihin soveltuvat työkalut tai dokumentointialustat tuottavat kuvailutietoa, kuten esimerkiksi alempana kohdassa 6 mainitut Arches ja iDAI.field.

Käytännössä kuvailuun on kolme vaihtoehtoa:

- Metadata sijoitetaan tiedoston sisään, esimerkiksi toimisto-ohjelmätiedostoissa tai kuvatiedostoissa. Tällöin metadata pysyy tiedon mukana mutta voi olla vaikeasti löydettävissä.

- Metadata sijoitetaan ulkoiseen tiedostoon, joko datatiedostokohtaisesti tai esimerkiksi koko kansiota koskevaan taulukkomuotoiseen tekstitiedostoon (CSV eli Comma Separated Values). Tällöin metadata ei välttämättä pysy itse datan mukana, mutta sen saavutettavuus ja löydettävyys on hyvä.
- Aineiston hallintaan käytetään järjestelmää, joka linkittää datan ja metadatan automaattisesti toisiinsa ja pitää huolen sekä datan eheydestä että metadatan löydettävyydestä. Tällaisissa ratkaisuissa datan käytettävyys osana tutkijan päivittäistä työntekoa on hankalaa, koska aineisto ei ole suoraan omalla koneella hyödynnettävissä.

Mikään näistä ratkaisuista ei ole täysin tyydyttävä.

Mikä tahansa näistä ratkaisuista on parempi kuin ei mikään näistä ratkaisuista.

6) TUTKIMUKSENAIKAINEN SÄILYTYS

Muista versionhallinta ja varmuuskopiointi. Tee niistä rutiini tutkimushankkeessasi. Jos kyseessä on useamman henkilön työllistävä hanke, sopikaa rutiinit ja kirjatkaa ylös aineistonhallintasuunnitelmaan, jotta ne ovat kaikkien tutkimukseen osallistuvien tarkistettavissa ja käytettävissä.

Instituutilla voi olla omat pilvipalveluratkaisunsa, kuten tämän hankkeen aikana käytössä ollut TeamDrive, joidenkin suomalaisten yliopistojen käyttämä Seafile tai esimerkiksi ownCloud. Osa näistä on asennettavissa myös instituutin omille palvelimille.

EUDAT-palveluun kuuluva B2DROP on yksi mahdollinen vaihtoehto, jonka käyttö on yksittäisillekin tutkijoille tai tutkimusryhmille mahdollinen ja EU-kansalaisille maksuton. Sen käyttö ei siis edellytä tutkijan kuulumista johonkin kyseisessä palvelussa olevaan tutkimusorganisaation (<https://www.eudat.eu/services/b2drop>). EUDAT-palveluun on mahdollista liittää myös B2SHARE, jossa tutkimusaineistoja voi julkaista ja jossa niille voi saada mm. DOI-tyyppisen pysyvän tunnusteen.

Toimiva vaihtoehto saattaa myös olla jonkinlainen aineistonhallintajärjestelmä, jonka avulla on mahdollista hallinnoida suuriakin määriä digitaalisia objekteja sekä mahdollisesti myös kuvata itse tutkimuskohteita. Tällaisia järjestelmiä ovat mm. Arches, Islandora, eXist-db ja iDAI.field.

AINEISTONHALLINTAJÄRJESTELMIÄ

iDAI.field on erityisesti arkeologiseen kenttätyöhön suunniteltu ohjelmisto, joka asennetaan pilveen ja jota käytetään mobiililaitteilla suoraan kentällä. iDAI.field on osa Saksan arkeologisen instituutin (DAI) aineistonhallinnan työkaluja (<https://idai.world/>).

Arches on erityisesti muistitieto-organisaatioiden käyttöön kehitetty järjestelmä, jossa on vahva tuki CIDOC-CRM-ontologialle ja sen tieteenalakohtaisille laajennuksille. CIDOC-CRM mahdollistaa esineiden, niitä kuvaavan dokumentaation ja itse tutkimusprosessinkin semanttisen mallinnuksen sellaisella tasolla, joka soveltuu myös monimutkaisten data-aineistojen tallentamiseen. The Getty Conservation Institute ja World Monuments Fund ovat Arches-järjestelmän kehittäjäorganisaatiot (<https://www.archesproject.org/>).

Islandora on erityisesti digitaalisen aineiston arkistointiin soveltuva järjestelmä, joka on parhaimmillaan suurten aineistojen hallinnassa ja julkaisussa (<https://islandora.ca/>).

eXist-db on xml-pohjainen tietokanta ja julkaisujärjestelmä, joka soveltuu erityisen hyvin esimerkiksi tekstin rakennetta kuvaavan TEI-pohjaisen tekstiaineiston julkaisemiseen.

7) VALMIS TUTKIMUSAINEISTO

DATA-ARKISTO ELI REPOSITORIO

Tässä vaiheessa (joulukuu 2020) suositellaan Zenodo-palvelua, joka on luotettava, laajasti eri tieteenaloilla käytetty, tuottaa pysyvän doi-tunnisteen ja hyväksyy eri tyyppistä aineistoa postereista datasetteihin. Kirjautuminen Zenodo.org -palveluun onnistuu ORCID-tunnuksella. Ks. ohjeet esim. <https://instruct-eric.eu/help/other/zenodo-upload-guidelines>.

YLEINEN, YKSILÖIMÄTÖN KÄYTTÖLUPA, ESIM. CREATIVE COMMONS -LISENSSI

Aineistoa julkaistaessa on sille valittava lisenssi, sillä ilman näkyvää lisenssiä aineisto ei ole periaatteessa kenenkään muun käytettävissä ilman erillisiä käyttö sopimuksia. Aineistojen julkaisussa on muistettava, että lisenssin voi itse määrittellä vain sellaiselle aineistolle, jonka itse omistaa. Toisin sanoen, jos jonkin tutkimushankkeen käyttöön on saatu kuvia vaikkapa paikallisilta museoviranomaisilta, ei niitä voi omavaltaisesti jakaa eteenpäin itse päättämällään lisenssillä eli käyttöluvalla. Tällaisen aineiston kohdalla on siis muistettava, että ne voivat olla julkaistavan tutkimusaineiston osana vain silloin, kun tähän on aineiston omistajan lupa joko alun perin tai julkaisua varten erikseen saatuna tai jos niillä on omistajan antama julkaisemisen mahdollistava lisenssi.

Tietyillä erityisaloilla on omia lisenssikäytänteitään. Esim. mikäli hankkeessa syntyy ohjelmakoodia, on sen julkaisemiseen suositeltavampaa käyttää ohjelmistoille tarkoitettuja lisenssejä. Näistä löytyy lisätietoja mm. <https://www.gnu.org/licenses/license-list.html>.

Tutkimusaineistojen julkaisussa suositellaan usein käytettävän Creative Commons eli CC-lisenssejä. CC-lisenssien käytöllä on myös laaja kansallinen ja EU:n laajuinen tuki. Apuna lisenssin valinnassa voi käyttää esim. Creative Commons-sivustolta löytyvää [lisenssiavainta](#). Se antaa myös lisenssitekstin ja selvittää niissä käytetyn termistön eri usealla eri kielellä.

Näiden CC-lisenssitermien lyhyet selitykset on koottu alle ja lisää tietoa CC-lisensseistä löytyy sivulta <https://creativecommons.fi/>. CC-termejä voi yhdistellä. Esimerkiksi. CC BY-NC tai CC BY-NC-ND ovat melko tavallisia.

CCO – ”Ei käyttörajoituksia.”

Tekijä ei aseta käyttörajoituksia aineistolle. Tämä soveltuu esim. tutkimushankkeita kuvaavan ylätasoinen metadatan jakeluun, jolloin näiden kuvailutietojen levittäminen ja hyödyntäminen on mahdollisimman helppoa. (Tämä ei tarkoita sitä, että tutkimus itsessään olisi CCo-lisensoitu.)

CC BY (ByAttribution eli Nimeä) – ”Saa käyttää, kunhan nimi mainitaan.”

Käytössä esim. metadatan lisenssivaihtoehtona useissa palveluissa. Esim. Helsingin yliopisto kannustaa jakamaan myös tutkimusaineistot tällä lisenssillä tieteen avoimuuden edistämiseksi. Mikäli tutkija haluaa, että muut tutkijat voivat hyödyntää hänen keräämiään aineistoja omissa tutkimuksissaan, on tämä käytännössä ainoa vaihtoehto; mutta tutkimusjulkaisujen kohdalla tämä ei välttämättä ole paras vaihtoehto.

CC NC (eli NonCommercial eli Ei kaupallinen) – ”Lupa annetaan ei-kaupalliseen käyttöön.”

Annat muiden kopioida, välittää, levittää ja esittää sinun tekijänoikeuksiisi kuuluvaa teosta sekä sen pohjalta tehtyjä muokattuja versioita teoksestasi vain epäkaupallisessa käytössä. Mikä on kaupallista ja mikä epäkaupallista on välillä epäselvää.

CC ND (NoDerivatives eli Ei muutoksia) – ”Aineistoa saa käyttää ja levittää ilman muokkausta.”

Annat muiden kopioida, välittää, levittää ja esittää sinun tekijänoikeuksiisi kuuluvaa alkuperäistä teosta, mutta et salli muokattujen versioiden tekemistä teoksesta.

CC SA (ShareAlike eli Jaa samoin) – ”Muokatut teokset on jaettava samalla lisenssillä.”

Annat muiden julkistaa omasta teoksestasi muokattuja teoksia vain samalla lisenssillä, jolla oma teoksesi on julkaistu.

YKSILÖITY KÄYTTÖLUPA

Käyttölupa ei ole sama asia kuin yllä mainitut lisenssit. Jos hankkeelle tai sen tutkijalle on annettu lupa käyttää ja julkaista kuvaa haluamallaan tavalla, on kyseessä osapuolten väliseen sopimukseen perustuva käyttölupa. CC-lisenssit taas määrittelevät, miten kuka tahansa asiaan voi suhtautua.

Aineistojen käyttöoikeuksia koskeva tieto on pidettävä tallessa ja aineistojen yhteydessä koko hankkeen ajan siten, että myös aineistojen julkaisuvaiheessa on mahdollista erottaa aineisto, jota ei voi julkaista.

PYSYVÄ TUNNISTE ELI PID

Pysyvä tunniste eli PID (Persistent Identifier) on yksiselitteinen ja ainutkertainen julkinen nimi, joka annetaan tutkimusaineistolle tai sen osalle. Aineistoihin liittyvät tunnisteet tulevat julkaisemisen yhteydessä, eikä niistä huolehtiminen ole yleensä tutkijan omalla vastuulla.⁴

On tärkeää huolehtia myös omasta näkyvyydestään aineiston julkaisun yhteydessä. Vahvassa nousussa näyttäisi on ORCID-tunniste, jonka tutkija voi luoda itselleen affiliaatioistaan riippumatta. Lisätietoja tunnisteesta löytyy sivulta <https://tutkijatunniste.fi/>.

Erilaisia PID-tyyppejä:

- URI eli Uniform Resource Identifier, joka ilmaisee tiedon
 - paikan (URL eli Uniform Resource Locator, joka esim. kertoo www-sivun http-protokollan avulla).
 - nimen (URN eli Uniform Resource Name, kuten julkaisuissa käytetty ISBN, joka itsessään ei sisällä ko. resurssin sijaintia).
- DOI eli Digital Object Identifier, joita käytetään julkaisuissa.
- Handle-tunnisteet, joita käytetään esimerkiksi julkaisuarkistoissa.
- [ORCID](#)-tunniste tutkijalle.
- [ISNI](#)-tunniste yhteisöille ja erityisesti muille henkilöille kuin tutkijoille.

⁴ Tunnisteista esim. Hakala J. 2018. ”URN:NBN ja muut toiminnalliset tunnistejärjestelmät”. *Tietolinja* 2. <http://urn.fi/URN:NBN:fi-fe2018093036991>. Avoimen tieteen koordinaatiossa tuotettu kansallinen suositus pysyvien tunnisteiden käytöstä: <https://doi.org/10.5281/zenodo.3560738>.

A. LIITE: INSTITUUTTIIEN AINEISTONHALLINTASUUNNITELMAN MALLI

Tähdellä * merkityt kohdat vastaavat pitkälti **tutkimusaineiston** README-tiedoston tietoja.

Tutkimusprojektin yleiset tiedot

- *Projektin nimi:
- *Tekijä eli vastuullinen tutkija:
- *Edellisen ORCID-tunniste:
- *Organisaatio(t) (instituutti + muut mahdolliset organisaatiot, jotka mukana projektissa):
- Rahoittaja(t) (+ rahoituspäätöksen numero, jos jo tiedossa):
- Projektin tiivis kuvaus (abstrakti):
- Tekijän yhteystiedot (sähköposti, puhelin):

Tutkimusaineiston yleiset tiedot

- Aineiston alkuperä (onko kyseessä tutkimusta varten kerättävä aineisto ja/tai tutkimuksessa syntyvä aineisto ja/tai uudelleen käytettävä aineisto):
- Aineistotyyppi (esim. analoginen, digitaalinen/numeerinen data, taulukko, kuva, teksti, mittaus, paikkatieto, muistiinpanot):
- Tiedostomuodot ja käytettävät ohjelmat:
- Arvio aineiston koosta:
- *Alueellinen kattavuus (voi olla tekstimuotoinen kuvaus tai standardi):
- *Ajallinen kattavuus (voi olla tekstimuotoinen kuvaus tai standardi):
- Jatkokäytön tarpeen arviointi eli kenelle tai millaiseen tutkimukseen aineistosta voi olla hyötyä tämän tutkimusprojektin jälkeen:
- Täydentykö aineisto tämän tutkimusprojektin jälkeen:

Eettiset ja oikeudelliset kysymykset – eli mahdolliset rajoitukset, jotka liittyvät aineiston jatkokäyttöön

A) Henkilötiedot

Sisältääkö aineisto henkilötietoja?

Ei

Kyllä, nimi tai yhteystiedot.

Kyllä, henkilötunnus, syntymäaika tai ikä.

Kyllä, sukupuoli.

Kyllä, ammatti, koulutus, työpaikka, opiskelupaikka tai koulu.

Kyllä, kotitalouden koostumus.

Kyllä, siviilisääty.

Kyllä, ajoneuvon rekisterinumero.

Kyllä, ainutkertaiset elämäntapahtumat, arkaluonteiset henkilötiedot (terveydentila, uskonnollinen tai poliittinen vakaumus, seksuaalinen suuntautuminen).

Jos vastasit mihinkään yllä olevista kysymyksistä ”kyllä”, aineistosta pitää laatia tieteellisen tutkimuksen tietosuojarekisteriseloste: Onko sellainen tehty? (Kyllä/Ei -> Laadi tietosuojaseloste oman tutkimusasi käytäntöjen mukaan).

Jos aineistossa on henkilötietoja, onko siitä saatavissa

- anonymisoitu aineisto (tunnisteeton aineisto eli aineistoa ei voida ”palauttaa” henkilötasolle)
- pseudonymisoitu aineisto (aineistosta on olemassa koodiavain, jonka avulla henkilöt voidaan tunnistaa -> henkilötietosuojan piirissä oleva aineisto)
- ei, aineistoa ei ole anonymisoitu eikä pseudonymisoitu (-> henkilötietosuojan piirissä oleva aineisto)
- B) Sopimukset ja lisenssit (Tarkista, että aineiston README-tiedostossa on samat tiedot)
- Käytätkö tutkimuksessa aineistoa, jonka käyttöä ja/tai myöhempää julkaisemista rajoittaa tekijänoikeus, lisenssit tai muut syyt? Jos vastaat kyllä, kerro millaisista rajoituksista on kysymys.
- Millä tavoin aineiston tekijänoikeudet, julkaisu-oikeudet ja omistajuus sovitaan (esim. kirjallinen sopimus, suullinen sopimus, kenen kanssa sopimus tehdään, missä kirjalliset sopimukset säilytetään)?
- Millä lisenssillä arvioit siirtäväsi aineiston data-arkistoon? (Ks. Creative Commons -lisenssiapu: <https://creativecommons.org/choose/?lang=fi>):

Tutkimusaineiston dokumentaatio ja kuvailu (metadata/metatieto)

- Millä tavoin varmistat, että aineisto on ymmärrettävää? (Kirjoita esim. näin ”README-tiedostossa annetaan aineiston perustiedot, joiden perusteella aineisto on ymmärrettävää.”)
- Tässä yhteydessä voi mainita myös aineiston ymmärrettävyyttä varmistavien muistiinpanojen tallennus ja mahdollinen julkaisu.
- *Kuvailussa käytetty metadatastandardi (esim. Dublin Core):
- Jos kuvailussa ei käytetä metadatastandardia, miten aineisto kuvaillaan:
- *Käytetyt sanastot (esim. Finto <https://finto.fi/fi/> + ne sanastot, joita käytät, esim. MAO/TAO tai PTO – Paikkatieto-ontologia jne; Getty Vocabulary, Art & Architecture Thesaurus <https://www.getty.edu/research/tools/vocabularies/aat/>):
- Jos kuvailussa ei käytetä valmiita sanastoja, miten käytettyjen termien sisällön ymmärrettävyys varmistetaan:

Tutkimusaineiston säilytys tutkimuksen aikana

- Missä ja miten aineisto säilytetään (digitaalinen aineisto: esim. oman tietokoneen kiintolevy, instituutin tietokone, verkkolevy, muistitikku, ulkoinen kiintolevy, instituutin pilvipalvelu, julkinen pilvipalvelu (Google Drive, Dropbox jne), data-arkisto, jne.; analoginen aineisto: esim. instituutin tilat, instituutin arkisto, tutkijan koti jne.):
- Kenellä on pääsy aineistoon ja miten pääsyä hallinnoidaan (koskee erit. projekteja ja arkaluonteista tietoa sisältävää tutkimusta):
- Miten digitaalisen aineiston luettavuus turvataan, kun käytetyt ohjelmistot vanhenevat:

Tutkimusaineiston säilytys ja julkaiseminen tutkimuksen päätyttyä

- Minkä osan aineistosta voi julkaista ja/tai avata?
- Milloin aineisto ja/tai sen kuvailutiedot avataan (eli sis. mahdollisen embargon)?
- Missä aineisto ja/tai sen kuvailutiedot avataan (esim. sama kuin pitkäaikaissäilytyksen data-arkisto)?
- Arvio aineiston pitkäaikaissäilytyksen kestosta (x vuotta)?
- Esim. Zenodo, EUDAT (tutkimuksen aikainen säilytys)

Tutkimusaineiston hallinnan vastuut ja resursointi

- Kuka vastaa tutkimusaineiston hallinnasta tutkimushankkeen ajan (esim. erityiset tehtävät, jotka eivät kuulu tutkimuksen arkipäivän rutiineihin)?
- Arvio ajallisen, tiedollisen/taidollisen ja rahallisen resursoinnin tarpeesta:
- Onko tämä huomioitu tutkimushankkeen suunnitelmassa ja rahoitushauissa/rahoituksessa

B. LIITE: README-TIEDOSTO

Tämän [AINEISTON NIMI] readme.txt tiedoston on luonut [VVVKKPV] [Sukunimi, Etunimi]

YLEISET TIEDOT

Aineiston:

*Nimeke: [esim. tutkimusaineisto artikkeliin/projektiin/julkaisuun "x"]

*Tekijä: [Sukunimi, Etunimi]

*Tekijän ORCID-tunniste:

Muu(t) tekijä(t): [Sukunimi, Etunimi; Sukunimi, Etunimi]

Muiden tekijöiden ORCID-tunnisteet:

Organisaatio(t): [esim. Suomen xxx-instituutti]

Alueellinen kattavuus: [voi olla tekstimuotoinen kuvaus tai standardi]

Ajallinen kattavuus: [voi olla tekstimuotoinen kuvaus tai standardi]

JAKAMINEN/SAAVUTETTAVUUS/PÄÄSY AINEISTOON

Lisenssit: [esim. Creative Commons]

Rajoitukset: [huomioi erityisesti, sisältääkö arkaluonteista aineistoa]

Aineistoon viittaamisen suositus: [nimi, jolla halutaan aineistoon viitattavan]

Linkki tähän aineistoon:

Linkki muuhun sijaintiin, jossa on avoin pääsy tähän aineistoon (kokonaan tai osaan siitä):

AINEISTO JA TIEDOSTOT: YLEISTÄ

Tiedostolista: [tiedostojen nimet, rakenne, lyhyt kuvaus tiedostoista]

Tiedostojen suhde toisiinsa, jos se on olennaista:

Lista muista lähteistä, jos aineisto on niistä peräisin:

METODIT

Kuvaus metodeista, joiden avulla aineisto on koottu tai kerätty mahdollisine linkkeineen:

Kuvaus metodeista, joilla aineisto on analysoitu:

Ohjelmistot ja laitteet, joita tarvitaan aineiston tulkintaan (muista ohjelmistojen versionumerot):

Standardit ja kalibroinnit, jos tarpeen mainita:

AINEISTOON LIITTYVÄ INFORMAATIO (Luo oma osio jokaiselle tiedostolle tai datasetille, jos se on tarpeen.)

Muuttujat, lyhenteet, mittayksiköt, symbolit, koodit, jne. (eli sellainen informaatio, jota tarvitaan aineiston perusymmärrykseen):

Metatietostandardit ja sanastot, joita on käytetty aineiston kuvailussa: