

Tietoarkiston palvelut

Arja Kuula-Luumi (Tietoarkisto)

Tieteelliset lehdet ja tutkimusdata –seminaari 1.2.2017

Tieteiden talo, Helsinki



TIETOARKISTO



TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE

Sisältö

- > Tiedejulkaisujen datapolitiikoista
- > Tietoarkisto
- > Tietoarkiston aineistopalvelut ja tuki



Tutkimusaineiston avaaminen tiedejulkaisussa

= Julkaisun tulosten taustalla oleva
aineisto avataan tulosten
tarkistamista ja tieteellistä
jatkokäyttöä varten



Yhä useammassa lehdessä

- Julkaisijoiden kanta vaihtelee merkittävästi paitsi aloittain, myös kustantajittain
- Artikkelin hyväksytään julkaistavaksi, mikäli tulosten taustalla oleva aineisto on muun tutkijakunnan käytettävissä artikkelin julkaisun jälkeen



Käytäntö

- Artikkeeliin on kirjattava tieto missä artikkelin aineisto on saatavilla.
- Julkaisijat suosittelevat aineistojen avaamista asianmukaisessa data-arkistossa
- Joissain tapauksissa aineiston voi tallentaa myös artikkelin lisätieto-osioon
- Aineistoihin täytyy viitata asianmukaisesti



Käytäntö

- Mikäli aineistoa ei voida avata esimerkiksi lainsäädännöllisistä tai tekijänoikeudellisista syistä, syy täytyy perustella
- Useimmiten kirjoittajien täytyy olla valmiita siinäkin tapauksessa luovuttamaan aineisto tarvittaessa tulosten tarkistusta varten



Avoimuutta vaativia julkaisijoita





EDITORIAL



**Sharing Clinical Trial Data — A Proposal from the International
Committee of Medical Journal Editors**

- Jatkossa ICMJE:n jäsenlehdissä hyväksyttäisiin julkaistavaksi vain sellaisten kliinisten tutkimusten artikkeleita, joiden tulosten taustalla oleva tutkimusaineisto avataan <http://www.icmje.org/news-and-editorials/M15-2928-PAP.pdf>
- Eettinen näkökulma



Datajulkaisut

- > Data paper (Datajulkaisu)
- > Metadatadokumentti, joka kuvailee yksityiskohtaisesti aineistotiedoston tai aineistosityn (ei analyysinä!)
- > Julkaistaan vertaisarvioituna artikkelina
 - Datajulkaisuja sisältyy seuraaviin joulaleihin: Nature Scientific Data, Gigascience, F1000 Research, ZooKeys, PhytoKeys



TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE



Tietoarkisto

Tietoarkisto on tutkimuksen ja opetuksen valtakunnallinen palveluinfrastruktuuri. Se arkistoi ja välittää sähköisiä tutkimusaineistoja tutkimukseen, opetukseen ja opiskeluun. Kaikki palvelut ovat maksuttomia.

Tietoarkisto edistää tieteen avoimuutta, läpinäkyvyyttä, tiedon kumuloitumista ja olemassa olevan tiedon tehokasta tutkimuskäyttöä. Tietoarkisto on ensimmäinen organisaatio Suomessa, jolle on myönnetty luotettavan pitkäaikaissäilyttämisen **Data Seal of Approval (DSA)** -sertifikaatti.

› [Lue lisää Tietoarkistosta](#)





Mitä Tietoarkistoon voi arkistoida? (1/2)

- > havaintomatriiseina tallennettavat tilastomatemattiset aineistot (tyypillisimpiä ovat kyselylomakkeisiin perustuvat aineistot)
- > haastattelujen, ryhmähaastattelujen ja erilaisten vuorovaikutustilanteiden litteraatit
- > tutkittavien kirjoittamat aineistot (päiväkirjat, temaattiset kirjoitukset ja vastaavat)
- > tutkijan tai tutkittavan ottamat valokuvat (ks. [oikeuksien siirrosta sopimisesta](#))



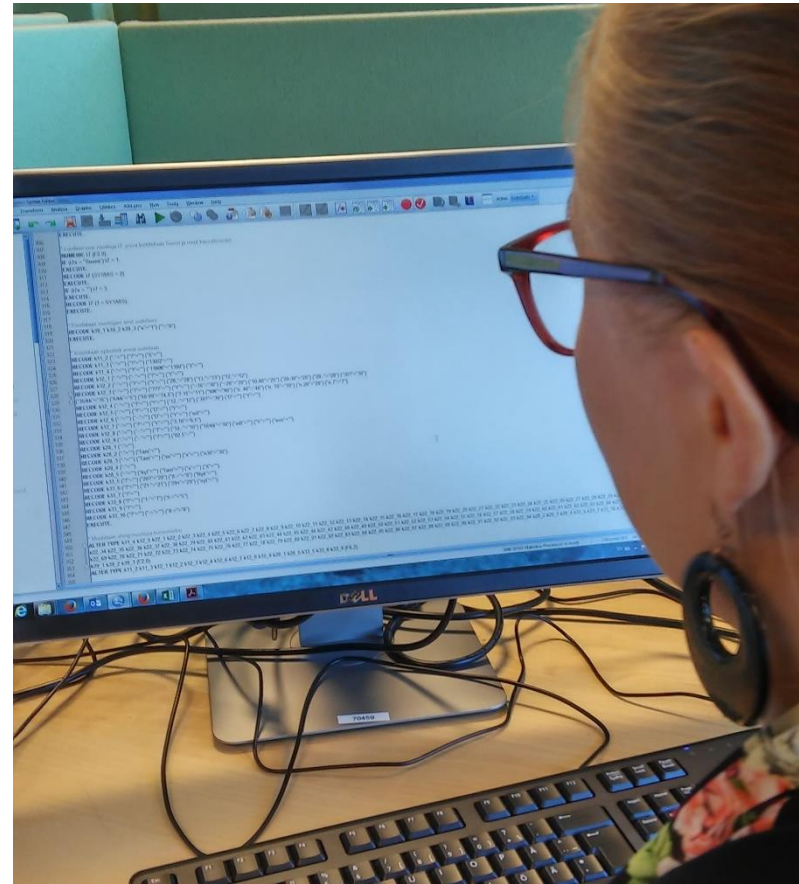
Mitä tietöarkistoon voi arkistoida? (2/2)

- > tutkimuksessa analysoidut lehtiartikkelit, lehtikuvat, kuvitukset, sarjakuvat (Tietoarkiston ja Kopioston välinen sopimus kattaa kotimaiset ja ulkomaiset teokset, mutta ei av-teoksia)
- > tutkijan digitoimat tai digikuvaamat analogiset arkistoaineistot, kun ne on digitoitu tai kuvattu tutkimuksessa analysoitavaksi ja niitä koskevat viitetiedot ovat riittävät (ks. [ohjeet](#))
- > **Tietoarkisto voisi opastaa rahoituksen saaneita tutkijoita jo projektin alusta saakka, jos tietäisimme, ketkä suunnittelevat palvelujemme käyttöä!**



Arkistointi

- > Arkistoitava aineisto käydään läpi huolella
 - Muuttujien tunnistus, muuttujien saamat arvot, puuttuvat tiedot
 - Sisäisen metadatan (tiedostojen nimet, aineistoyksikköjen taustatiedot) konsistenssi
 - Anonymisoinnin tarkistus
- > Aineiston kuvailu
- > Kuvailutietojen julkaiseminen luetteloissa: suomeksi ja englanniksi





TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE

Jatkokäytön tuki – Aila



TIETOARKISTO

Aila

Aineistot Haku Ohjeet

In English

Hei vieras!

Kirjaudu

Rekisteröidy

/ Etusivu

1. Kirjaudu Ailaan
2. Valitse sopiva aineisto
3. Lataa aineisto nopeasti ja kätevästi



Lataa dataa Ailasta

Ailasta löydät Tietoarkistoon tallennetut tutkimusaineistot sekä niiden kattavat suomen- ja englanninkieliset kuvailutiedot. Aineistojen selaaminen, hakeminen ja kuvailutietojen lukeminen on täysin vapaata. Osa aineistoista on myös avoimesti ladattavissa.

Rekisteröitymällä asiakkaaksi pääset maksutta lataamaan käyttöehdoiltaan rajattuja aineistoja. Ailassa on yli 1200 aineistoa.

Rekisteröityminen tapahtuu yliopistojen ja ammattikorkeakoulujen Haka-tunnuksilla. Voit myös hakea käyttäjätunnusta asiakaspalvelustamme.

Huoltoikkuna tiistaisin kello 7–10

Kirjaudu

Uusi käyttäjä? Rekisteröidy.

Ajankohtaista Ailassa

- › Tietoarkisto palvelee myös kesällä
- › Aila-seminaarin esiintyjälista on valmis
- › Muutoksia Ailan käyttäjäprofileihin

Aineistoluettelo

Selaa, hae ja lataa aineistoja

Ailan ohjeet

Lue Ailan käyttöohje

Ailan käyttöehdot

Lue Ailan käyttöehdot

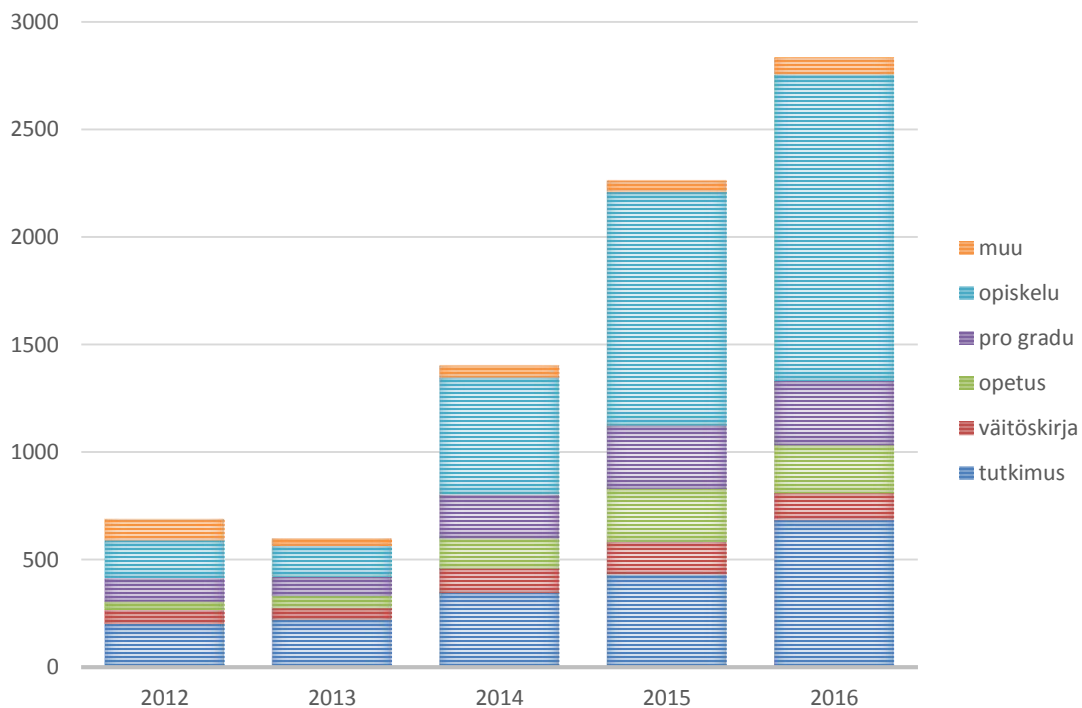


Palveluportaali Aila

- > Noin 1300 aineistoa
- > Data saatavilla arkistointisopimuksissa määriteltyjen ehtojen mukaisesti
 - (A) vapaasti käytettävissä (58),
 - (B) tutkimukseen, opetukseen ja opiskeluun (924),
 - (C) vain tutkimukseen graduista alkaen (166) ja
 - (D) vain luovuttajan luvalla (104).



JATKOKÄYTTÖÖN TOIMITETTUJEN AINEISTOJEN LUKUMÄÄRÄ KÄYTTÖTARKOITUKSEN MUKAAN 2012-2016



- ✓ Aineistoja kaikkiaan 1300
- ✓ Aineistojen käyttöaste 84%

Englanninkielisestä
aineistoluettelosta ladattu
noin 25% kaikista
aineistolatauksista



Asiakkaat eri puolilta maailmaa





Arkistointi edistää ansioitumista

- > Aineistojen jakaminen mainitaan tutkijan CV:ssä.
- > Aineistoon on viitattava kuten julkaisuun viitataan.
- > Aineistojen jatkokäyttö lisää viittauksia myös tutkimusjulkaisuihin.
- > Aineistojen jakaminen lisää yhteistyötä (Suomessa ja kansainvälisesti).



Horizon 2020: DMP should include information on

- > the handling of research data during and after the end of the project
- > what data will be collected, processed and/or generated
- > which methodology and standards will be applied
- > whether data will be shared/made open access and
- > how data will be curated and preserved (including after the end of the project)



Horizon 2020: FAIR periaatteet

- > F = findable
 - Hyvä ja kattava metadata, pysyvät tunnisteet
- > A = accessible
 - Aineiston metadata julkaistaan avoimissa dataluetteloissa ja se on aina löydettävissä URNinsa perusteella
 - Tiedeyhteisö saa datan käyttöönsä
- > I = interoperable
 - Metadatan tuottamiseen käytetään selkeää ja formaalia kieltä
 - Suosittava yhteisiä sanastoja
 - Referenssit muihin data standardeihin
- > R = reusable
 - Metadata sisältää riittävästi attribuutteja
 - Provenienssitiedot
 - Käyttöehdot (lisenssit)



Avointa metadataa

- > Kaikki Tietoarkiston aineistokuvailut DDI Codebook XML-formaatissa.
 - Rikas metadata avoimesti haettavissa ja selattavissa – data rajoitetummin saatavilla
 - Suppea versio kuvailudatasta kerättävissä Dublin Core- tai EAD2-formaateissa ([ks. artikkeli KUHAsta](#))
 - Saatavilla myös muissa luetteloissa/ palveluissa



FINNA

cessda





Aineistohallinnan käsikirja

1 Miksi aineistohallintaa ja jatkokäyttöä?

- Osa hyvää tieteellistä käytäntöä
- Rahoittajat edellyttävät aineistohallintaa
- Jatkokäytön ja avoimen saatavuuden hyödyt
- Tutkimusaineiston elinkaari

2 Aineistohallinnan suunnittelu

- Aineistohallintasuunnitelma
- 1. Aineisto
- 2. Oikeudet
- 3. Tietosuoja ja tietoturva
- 4. Tiedostoformaattit ja ohjelmistot
- 5. Aineiston käsittelyn ja sisällön kuvaus
- 6. Elinkaari

3 Sopimukset ja oikeudet

- Tekijänoikeus
- Tutkijoiden välisistä sopimuksista
- Oikeuksien siirtymisestä sopiminen
- Tekijäyys tieteellisissä julkaisuissa ja tutkimusaineistoissa

4 Tutkittavien informointi

- OSA I: Informoinnin merkitys
- OSA II: Informoinnin sisältö, muoto ja tutkittavan suostumus
- OSA III: Informoinnin osa-alueet
- Osa IV: Esimerkkejä informoinneista

5 Tiedostoformaattit ja ohjelmistot

- Tiedostojen konvertointi
- Tutkimusaineiston digitointi
- Havaintomatriisit ja tilasto-ohjelmistot
- Tekstiaineistot
- Kuva-aineistot
- Ääni- ja äänikuvatallenteet

6 Kvantitatiivisen datan käsittely

- Havaintomatriisin tallennus
- Muuttujien nimet ja selitteet
- Muuttujan arvot ja niiden selitteet
- Muuttujien muunnokset
- Puuttuva tieto
- Painomuuttujat
- Syntaksin käyttö

7 Kvalitatiivisen datan käsittely

- Litterointi
- Aineiston sisältämien datatiedostojen organisointi
- Datatiedostojen nimeäminen
- Taustatietojen kirjaamistavat
- Lehtiaineistot

8 Tunnisteellisuus ja anonymisointi

- Mitä on henkilötieto?
- Milloin tieto on tunnisteellista?
- Milloin tieto on anonymiä?
- Tunnisteellisten aineistojen käsittely
- Anonymisoinnin lähtökohdat
- Kvantitatiivisen aineiston anonymisointi
- Kvalitatiivisen aineiston anonymisointi

9 Aineiston kuvailu ja metadata

- Metadatan tallennus
- Tutkimuksen toteutuksen kuvailu
- Aineistonkeruuinstrumentti
- Tiedostojen kuvailu
- Muuttujien kuvailu
- Saatavuustiedot
- Kontekstitiedot ja paradata
- Metadataformaattit

10 Fyysinen säilytys

- Tiedostojen nimeäminen ja hallinta
- Varmuuskopiointi
- Migraatio ja virkistäminen
- Tallennusmediat
- Tietoturva
- Aineiston hävittäminen
- Paperiaineiston säilytys

11 Malliesimerkit ja sanastot

Sisällys





TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE

Kiitos!

[asiakaspalvelu.fsd @ uta.fi](mailto:asiakaspalvelu.fsd@uta.fi)