



**OHJE TIETEELLISILLE  
JULKAISUKANAVILLE  
VASTUULLISEN AINEISTO- JA  
DATAPOLITIIKAN LAATIMISEKSI**

## VASTUULLISEN TIETEEN JULKAISUSARJA

Vastuullisen tieteen julkaisusarjassa julkaistaan julistuksia, linjauksia, selvityksiä, suosituksia ja muita asiakirjoja, jotka liittyvät tieteen ja tutkimuksen avoimuuteen, vastuullisuuteen, saavutettavuuteen, tiedeviestintään sekä tieteen ja yhteiskunnan vuorovaikutukseen. Julkaisusarja ei ole tieteellinen vertaisarvioitu julkaisu. Julkaisusarjaa julkaisevat Tieteellisten seurain valtuuskunta sekä Tiedonjulkistamisen neuvottelukunta.

Ohje tieteellisille julkaisukanaville vastuullisen aineisto- ja datapolitiikan laatimiseksi

Vastuullisen tieteen julkaisusarja 4/2020  
Publikationsserie för ansvarsfull forskning 4/2020  
Responsible Research Series 4/2020

Tekijät: Jyrki Ilva, Susanna Nykyri, Henriikka Mustajoki, Jessica Parland-von Essen ja Sami Syrjämäki

Julkaisija: Tiedonjulkistamisen neuvottelukunta ja Tieteellisten seurain valtuuskunta

Helsinki, 2020

ISSN 2670-062X  
ISBN: 978-952-5995-22-0  
DOI: <https://doi.org/10.23847/isbn.9789525995220>



**OHJE TIETEELLISILLE  
JULKAISUKANAVILLE VASTUULLISEN  
AINEISTO- JA DATAPOLITIIKAN  
LAATIMISEKSI**



**V**astuullisuus, läpinäkyvyys, avoimuus ja toistettavuus ovat tieteen kulmakiviä. Kestävällä tavalla toteutettu avoin tiede ja vastuullinen datan hallinta edistävät näiden tavoitteiden toteutumista.<sup>1</sup>

Myös yhä useampi rahoittaja ja tiedepoliittinen toimija laatii linjauksia, jotka ohjaavat tutkimusyhteisöä kohti avoimen tieteen käytäntöjä.

Tutkimus perustuu usein dataan tai aineistoon, jota ei ole mahdollista julkaista varsinaisessa tutkimusartikkelissa. Riippumatta siitä, onko data julkaistavissa tai ei, ovat julkaisukanavien vastuullinen datapolitiikka ja julkaisijoiden tutkijataidot keskeisessä roolissa, kun pyritään toteuttamaan yllä mainittuja tavoitteita.<sup>2</sup>

- Tämä data- ja aineistopolitiikkasuositus tiedejulkaisijoille on laadittu tukemaan julkaisukanavia ja julkaisijoita, jotta he voisivat vastata kansainvälisten ja kansallisten sekä yksittäisten tutkimusrahoittajien ja muihin avoimen tieteen linjausten haasteisiin. Ohjeistus ohjaa käytäntöihin, jotka ovat yhteensopivia näiden linjausten ja tutkimustyön yleisten tavoitteiden kanssa datan ja aineistojen avaamisen

1 Tieteen termipankki määrittelee käsitteet seuraavasti: Tutkimusaineisto: tutkijan tuottama tai tutkimusprosessin aikana käyttämä resurssi eli digitaalisessa, analogisessa tai fyysisessä muodossa oleva tieteellisen tutkimuksen perusaineisto tai perusaineistosta jalostettu aineisto, johon tutkimuksen löydökset ja tulokset perustuvat. Data: digitaalinen aineisto, johon ei liity tulkintaa (<https://tieteentermipankki.fi/wiki/Termipankki:Etusivu>)

2 Center for Open Science (Charlottesville, VA, USA) on luokitellut jornaalien avoimen datan politiikkaa tasoluokkiin välillä 0–3 (0 tarkoittaa rohkaisemista joihinkin käytäntöihin ja toisaalta pidättäytymistä kannanotoissa toisiin. Taso 3 puolestaan vaatii kirjoittajilta vahvaan avoimuuteen sitoutumista. Taso 3 saattaa vielä tois- taiseksi olla liian vaativa yleiseksi periaatteeksi. Se vie liikaa resursseja kirjoittajilta ja kustantajilta, mutta taso 0 on kansallisten ja kansainvälisten linjauksien valossa liian heikko. Vahvempi datapolitiikka ja nollatasoa tarkempi ohjeistus säästää kaikkien osapuolten aikaa, koska näin vältetään väärinkäsityksiltä, ja kun standardeista pidetään heti huolta, datan tai aineiston käyttäminen myöhemmin uudestaan onnistuu helpommin (<https://science.sciencemag.org/content/348/6242/1422>).

eettiset kysymykset huomioiden.<sup>3</sup> Kannustamme sekä tutkijoita että kustantajia vaalimaan myös FAIR-data-periaatteita.<sup>4</sup>

Koska kyseessä on yleisen tason ohje, tulee ohjetta sovellettaessa huomioida eri alojen ja tutkimusmenetelmien erityispiirteet. Erityisesti sensitiiviseksi luokiteltavan datan ja aineistojen avaamiseen liittyvissä kysymyksissä on eroja, jotka asettavat rajoituksia sen avaamiselle. Tällaisia eroja on tyypillisesti esimerkiksi laadullisen tutkimuksen ja kvantitatiivisen tutkimuksen aineistojen välillä, koska laadullisten syvähaastattelujen anonymisointi on usein haasteellisempaa kuin kvantitatiivisten aineistojen, ja toisinaan se on käytännössä mahdotonta.

Avoimuuden asteet vaihtelevat. Aineiston täydelliseen avaamiseen saattaa liittyä eettisiä, teknisiä tai resursseihin liittyviä esteitä, mutta jos aineistoa tai edes sen osia ei ole jostain syystä mahdollista saattaa minkään tutkijan itsensä tai tutkimusryhmän ulkopuolisen tahon arvioitavaksi edes erityisluvalla, ei tiedeyhteisö voi arvioida aineiston perusteella tehtyjen johtopäätösten osuvuutta tai aineiston laatua.

On aina parempi, että julkaisukanava laatii ja julkaisee oman aineistopolitiikkansa kuin jättää sen heitteille, vaikka julkaisukanava ei ottaisi voimakkaasti kantaa suuntaan tai toiseen. Julkinen politiikka selvittää julkaisukanavassa julkaistujen tutkimusten aineistojen ja datan saatavuutta, käyttömahdollisuuksia ja käytön rajoituksia tutkimusyhteisölle ja muille kiinnostuneille tahoille, ja ohjaa tutkijoita huomioimaan oleelliset seikat hyvissä ajoin.

Tämän ohjeen lähtökohta on Journal Research Data Policies (JoRD) -projektissa tieteellisiä lehtiä varten laadittu avoimen datan politiikan malli (2014). Journal Research Data -projektissa kartoitettiin julkaisujen olemassa olevia datalinjauksia ja haastateltiin eri sidosryhmiä, kuten eri alojen tutkijoita ja kustantajia. Projektin tuloksena syntyi tieteellisille lehdille

---

3 Linjauksista: katso esimerkiksi DORA-julistuksen osuus kustantajille (<https://sfdora.org/read/fi/>) ja Avoin tiede ja data -hankkeen linjaukset (<http://www.unifi.fi/toiminta/hankkeet/avoin-tiede/>). EU:n Horizon 2020 Guidelines for FAIR data management ([http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf))

4 FAIR-data periaatteet: <https://www.fairdata.fi/en/why-fairdata/fair-principles/>. Katso myös EU:n Horizon 2020 Guidelines for FAIR data management (linkki viitteessä 3)

tarkoitettu datapolitiikan kehikko, jossa kerrotaan, mihin seikkoihin yksittäisen lehden tulisi kiinnittää huomiota ja ottaa kantaa.<sup>5</sup> Alla oleva malli ei ole täsmälleen sama kuin JorRDin. Olemme pyrkineet kehittämään sitä viimeisimmän kehityksen ja erityisesti kotimaisen julkaisukentän huomioiden.

Lisätietoa:

Suositukset säilytys- ja siirtokelpoisille formaateille sekä muuta ohjeistusta koskien pitkäaikaissäilytystä: <http://digitalpreservation.fi/specifications>.

---

5 Sturges, Paul and Bamkin, Marianne and Anders, Jane H.S. and Hubbard, Bill and Hussain, Azhar and Heeley, Melanie (2014) Research data sharing: developing a stakeholder-driven model for journal policies. Journal of the Association for Information Science and Technology, "Research data sharing: Developing a stakeholder-driven model for journal policies", <https://doi.org/10.1002/asi.23336> Avoin vertaisarvioitu versio löytyy osoitteesta: <http://eprints.nottingham.ac.uk/3185/> ja projektin blogi osoitteesta: <https://jordproject.wordpress.com/>.

## VASTUULLISEN DATA- JA AINEISTOPOLITIIKAN MALLI TIEDEJULKAISULLE

Malli sisältää seitsemäntoista kohtaa, jotka tieteellisen julkaisijan olisi hyvä huomioida omassa data- ja aineistopolitiikassaan. Useisiin kohtiin on myös annettu sisällöllinen suositus.

### 1. Yleinen kommentti avoimen datan ja aineiston hyödyistä

Perustellaan datapolitiikan linjaukset sen hyödyillä.

Esimerkiksi: Tutkimuksessa käytetyn datan ja muun aineiston vastuullinen avoin jakaminen kohentaa tutkijayhteisön yhteistyötä, tehostaa tulosten arviointia, kannustaa kokeiden toistamiseen ja aineiston käyttämiseen uusissa tutkimuksissa ja edesauttaa sen yhteiskunnallista hyödyntämistä. Tutkija hyöttyy avoimuudesta saaden enemmän huomiota ja viittauksia työllensä, ja tutkimuksen laatu paranee. Jos aineistoa ei voi esimerkiksi sensitiivisyydestä johtuen avata, tulee aineiston keskeisten metatietojen löydettävyydestä silti huolehtia.

Yleisen tason ohje tutkijalle löytyy Vastuullinen tiede -sivulta: <https://www.vastuullinentiede.fi/fi/jatkokaytto/tutkijan-muistilista-tutkimusdatan-julkaisemiseen>

### 2. Poliittikadokumentin omistaja

Kerrotaan, minkä julkaisijan/julkaisun/tieteellisen seuran datapolitiikka tämä on.

Esimerkiksi: Tämä on "Julkaisun Tiedettä ja Tutkimusta" data- ja aineistopolitiikka.

### 3. Aineistojen ja datan alkuperäiset julkaisijat

Muistutetaan, että hyvän tieteellisen käytännön mukaisesti julkaisussa tulee aina mainita tutkimuksessa käytettyjen tietovarantojen tai aineistojen alkuperäiset julkaisijat.

Suositus: Julkaisun kirjoittajien tulee nimetä tutkimuksessa käytetyt aineistot ja niiden alkuperäiset julkaisijat.

#### 4. Poikkeustapaukset

Annetuista linjauksista tulee voida poiketa hyvistä perustelluista syistä. Syiden arvioimisprosessi tulee kuvata esim. "toimitus arvioi kokouksessaan poikkeustapauksen perusteet". Hyvä peruste voi olla esimerkiksi datan julkaisemisesta aiheutuvat vakavat haitat tutkimuksen kohteelle tai muulle toimijalle tai ettei sensitiivisen aineiston anonymisointi ole mahdollista kohtuullisin resurssein. Kirjoittajien olisi hyvä esittää perusteet aineiston avaamatta jättämiselle jo datan hallintasuunnitelmassa, mutta jos asiaa ei ole huomioitu siinä, perusteet tulisi esittää tarjotun artikkelin saatteessa.

Esimerkiksi: Annetuista linjauksista voi poiketa vain perustellusti. Julkaisun toimitus arvioi kokouksessaan poikkeustapauksen perusteet. Poikkeussyitä voivat olla esimerkiksi datan julkaisemisesta aiheutuvat vakavat haitat tutkimuksen kohteelle tai muulle toimijalle. Kirjoittajien tulee esittää perusteet suosituksista poikkeamiselle.

#### 5. Kannanotto datan säilytyspalveluun tallentamisen pakollisuudesta tai vapaaehtoisuudesta

Ohjeistetaan, kuinka datan tallentamisesta tulee huolehtia luotettavasti ja mahdollisimman avoimesti.

Suositus: Artikkelin kirjoittajien tulee tallentaa data tai muu tutkimusaineisto luotettavaan säilytyspalveluun tai varmistaa toistettavuus ja datan todennettavuus muilla tavoilla tutkimusalojen periaatteiden mukaisesti huomioiden eettiset ja juridiset rajoitteet.

#### 6. Datan ja aineiston säilytyspalvelut

Mainitaan suositeltavat datan ja aineiston säilytyspalvelut tai niiden valinnalle asetetaan selkeät kriteerit.



## 7. Datan tai aineiston liittäminen tai linkittäminen artikkeliin

Ohjeistetaan, kuinka data tai aineisto liitetään tai linkitetään julkaistavaan artikkeliin.

Suositus: Data tai aineisto tulee joko liittää artikkeliin tai linkittää artikkelissa muualla sijaitsevan aineiston kuvailusivuun (landing page).

## 8. Kannanotto linkittämisen muotoon

Kerrotaan, kuinka ja missä muodossa datalinkit annetaan.

Suositus: Linkitys tulee tehdä linkin pysyvyys ja uudelleenohjautuvuus varmistaen DOI tai URN -käytännöillä. Kansallinen ohjeistus: Data citation roadmap for Finland (<http://www.doria.fi/handle/10024/153328>)

## 9. Julkaisussa liitetiedostoina olevat data ja aineistot

Mikäli datan ja aineiston voi liittää julkaisuun liitetiedostoina, on huomioitava riittävä dokumentaatio ymmärrettävyyden takaamiseksi.

Jos datan ja aineistojen hyväksyttävälle tyypeille (kuva, taulukko, AV-aineisto yms.) ja formaatille (esim. avi, jpeg) on rajoituksia, hyväksyttävät tyypit tai kriteerit niiden valinnalle on mainittava.

Suositus: Jos liität julkaisuun dataa tai aineistoja tiedostoina, käytä [julkaisun suosittamia formaatteja]. Liitä mukaan myös riittävät metatiedot ja readme-tiedosto.

## 10. Laajemmasta datavarannosta valikoidut osat

Ohjeistetaan, kuinka mahdollisesti laajemmasta datavarannosta valikoitu osa liitetään artikkeliin.

Suositus: Mikäli julkaistu tutkimus perustuu suuremman datavarannon osaan, suositellaan joko kyseisen osan erillistä tallentamista, kyseisen datavarannon osan saattamista avoimeksi tai ainakin keskeisten osien tallentamista

toistettavuuden ja uudelleenkäytön mahdollistamiseksi.

### 11. Tutkimusdataan viittaaminen

Dataan tulee suhtautua samoin kuin muihin tutkimusjulkaisuihin. Datapolitiikassa ohjeistetaan, kuinka dataan tulisi viitata.

Suositus: Viittauksissa tulee käyttää kansallisen suosituksen mukaista tyyliä. Viittausten pakolliset ja täydentävät elementit on lueteltu sivulla 12: Tracing Data. Data Citation Roadmap for Finland (<http://www.doria.fi/handle/10024/153328>)

### 12. Vertaisarviointi, data ja aineistot

Otetaan kantaa datan ja aineiston saatavuuteen vertaisarviointiprosessissa.

Suositus: Vertaisarvioijan tulee päästä tarkastelemaan dataa ja aineistoja. Datan ja aineiston hallintasuunnitelma (DMP) tulee liittää käsikirjoitukseen.

### 13. Datan avaamisen embargo

Otetaan kantaa datan avaamisen mahdolliseen embargoon.

Suositus: Data avataan välittömästi, kun tutkimus julkaistaan. Vähintään keskeisten metatietojen on oltava julkaisuhetkellä avoimesti saatavilla.

### 14. Tutkimusetiikka ja juridiikka

Otetaan kantaa datan avaamisen eettisten ja juridisten kysymysten huomioimiseen.

Euroopan komission ohjeet datan suojaamisesta tietosuojasetukset huomioiden: [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/ethics/h2020\\_hi\\_ethics-data-protection\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-data-protection_en.pdf)

Suositus: Datan ja aineiston avaamisessa tulee huomioida siihen liittyvät eettiset ja juridiset kysymykset. Tutkijan tulee seurata Tutkimuseettisen neuvottelukunnan ohjeita (<https://www.tenk.fi/fi/eettinen-ennakkoarviointi-ihmistieteissa>)

## 15. Sensitiivisen datan ja aineistojen jakaminen

Julkaisu antaa ohjeet, kuinka yksittäiset tutkijat voivat tiedustella sensitiivisen datan saamista käyttöönsä. Julkaisu voi arvioida, onko ohjeet parempi antaa artikkeleiden yhteydessä vai muualla, esim. datapolitiikan linjauksissa.

Suositus: Jakajatahon yhteystiedot tulee antaa sensitiivisen datan jakamisen mahdollistamiseksi muiden tutkijoiden käyttöön.

## 16. Datan tai aineiston käyttöoikeudet

Selvitetään datan tai aineiston omistus- ja käyttöoikeudet. Kirjoittajien tulee perustella käyttöoikeuksien rajoitukset.

Suositus: Kirjoittajien tulee selvittää aineiston, datan tai niiden eri osien IPR-status eli antaa ohje niiden käyttöoikeuksista kirjoittajien, heidän instituutioidensa ja rahoittajien välillä. Jos dataa saa käyttää uudelleen, tulee kirjoittajien ilmoittaa käytön ehdot. Asian selkiyttämiseksi data tulisi lisensoida tai kirjoittajien tulee viitata muualta hankitun datan sopimukseen.

## 17. Datan analyysia koskevat tiedot

Kerrotaan, mitä tietoja datan analysointiin käytetyistä välineistä tulisi antaa.

Suositus: Datan analyysissa käytetyt ohjelmistot tulee mainita ja käytetyt koodit tulee kuvata.



Tieteellisten seurain  
valtuuskunta



TIEDONJULKISTAMISEN NEUVOTTELUKUNTA  
DELEGATIONEN FOR INFORMATIONSPRÖNING