

Tutkimusaineistojen ja - menetelmien avoimen saatavuuden linjaus

I. Tutkimusdatan avoimen saatavuuden osalinjaus

Avoimen tieteen kevätpäivät 13.5.2020

Pekka Orponen

Avoimen datan asiantuntijaryhmä, pj

Taustaa: valmisteluprosessi

- Lähetekeskustelu Avoimen tieteen strategiaryhmä 15.3.2019
- Valmisteluryhmä Pekka Orponen pj, Heidi Laine siht (02-10/2019)/Nina Järviö siht (11/2019-03/2020), 34 muuta jäsentä
- 10 etäkokousta 2019, 4 etäkokousta 2020, lähitapaamiset ja palautekeskustelut AT työpajapäivät 1.10.2019, 17.3.2020
- Lähtökohtana kotim ja kv taustamateriaalit, näistä ensin 9 ”avoimen datan pilaria” ja sitten 5 yleistä strategista periaatetta sekä 7 konkreettista tavoitetta vuosille 2021-2025
- Pilarien, periaatteiden ja tavoitteiden hahmottaminen laajan ryhmän kesken, sen jälkeen kullakin pilarilla (2019) ja tavoitteella (2020) oma valmistelutiimi, palautekeskustelut laajassa ryhmässä
- Tavoitteiden valmistelun pj:t: 1. Pekka Orponen (Aalto), 2. Susanna Nykyri (TUNI), 3. Ville Tenhunen (HY), 4. Mari Elisa Kuusniemi (HY), 5. MEK & Maria Rehbinder (HY), 6. Miki Kallio (OY), 7. Pekka Orponen (Aalto)
- Linjausluonnos V1.0 käsitelty AT ohjausryhmässä 23.4.2020
- AT ohjausryhmän palautteen perusteella V2.0 valmisteilla, tavoite toimittaa AT ohjausryhmän kokoukseen 11.6.2020 ja sen jälkeen lausuntokierrokselle

Taustaa: valmisteluryhmä 01/2020

Edgren-Henrichson	Nina	Svenska litteratursällskapet	Orponen	Pekka	Aalto & FCRD
Gustavsson	Pamela	Svenska litteratursällskapet	Päällysaho	Seliina	SeAMK
Järviö	Nina	TSV	Pursula	Antti	CSC
Karakainen	Meri-Tuulia	UTU	Rehbinder	Maria	Aalto-yliopisto
Kallio	Miki	OY	Riipinen	Mari	Turun yliopisto
Kantola	Anu	Luke	Riukulehto	Sulevi	HY Ruralia-instituutti
Karlsson	Liisa	HU	Rosti	Tomi	UEF
Kaurahalme	Olli-Pekka	Turun yliopisto	Ruuskanen	Matti	Turun yliopisto
Koiranen	Joona	Metropolia Amk	Saari	Toni	UEF
Kuusniemi	Mari Elisa	HY/Tuuli	Salmi	Jussi	UTU
Laine	Heidi	HY	Salminen	Nina-Mari	LUKE
Linnovaara	Kristina	SLS	Seppälä	Hannele	Kansall. koul. arv. keskus
Louhimo	Riku	Työterveyslaitos	Seppänen	Janne	Jyväskylän yliopisto
Marttila	Juuso	JY	Sipola	Tiina	OY
Naskali	Juhani	Turun yliopisto	Sunikka	Anne	Aalto-yliopisto
Neuvonen	Anssi	VTT	Suorsa	Anna	OY
Niku	Maria	SKS	Tenhunen	Ville	HY/Tike
Nykyri	Susanna	TAU	Tuomi-Nikula	Antti	THL

Taustaa: valmistelutyön lähtökohdat

- Aiemmat kotimaiset strategiadokumentit
 - ATT-hankkeen linjaukset <https://avointiede.fi/fi/linjauksia/att-hankkeen-keskeiset-linjaukset>
 - UNIFI Avoimen datan tiekartta <http://www.doria.fi/handle/10024/157394>
- Keskeiset kansainväliset dokumentit, esim.
 - Turning FAIR Into Reality, EC/DG Research 2018 <https://doi.org/10.2777/1524>
 - Prompting an EOSC in Practice, EC/DG Research 2018 <https://doi.org/10.2777/112658>
 - LERU Open Science Roadmap, LERU 2018 <https://www.leru.org/publications/open-science-and-its-role-in-universities-a-roadmap-for-cultural-change>
- Muu aihepiirin kirjallisuus ja ryhmän jäsenten ammatillinen asiantuntemus

Avoimen saatavuuden linjaus: päämäärä

- Tutkimusaineistot ja -menetelmät ovat niin avoimia kuin mahdollista ja niin suljettuja kuin välttämätöntä.
- Tietoaineistoja hallitaan tarkoituksenmukaisesti FAIR-periaatteita tavoitellen.
- Tutkimusmenetelmät ja -aineistot, mukaan lukien tutkimusdata, tunnistetaan itsenäisiksi tutkimustuotoksiksi.

Tutkimusdatan saatavuuden osalinjaus: käsitteitä (1/2)

- *Tutkimusaineisto*: Digitaalisessa, analogisessa tai fyysisessä muodossa oleva tieteellisen tutkimuksen perusaineisto, tai tällaisesta jalostettu tietoaaineisto, johon tutkimuksen löydökset tai tulokset perustuvat.
- *Tutkimusdata (tutkimuksen tietoaaineisto)*: Tietoaaineisto, joka on kerätty, havaittu, mitattu tai luotu tutkimustulosten todentamiseksi. *Tämä osalinjaus koskee ensisijaisesti digitaalista tutkimusdataa sekä muunlaisten tutkimusaineistojen digitaalisia kuvailutietoja (metadata).*
- Tutkimusdatan kolme päätyyppiä:
 - *Raakadata*: jalostamaton, esimerkiksi suoraan jonkin mittalaitteen tuottama data
 - *Tulosdata*: tutkimuskäyttöön jalostettu ja dokumentoitu data, joka voi kuitenkin vielä kumuloitua tai muuttua
 - *Datatuotos*: arkistointikelpoinen datakokonaisuus, joka vastaa johonkin tutkimuskysymykseen

Tutkimusdatan saatavuuden osalinjaus: käsitteitä (2/2)

Tutkimusdatan hyvä t. vastuullinen hallinta: Tutkimuksen tietoaineistojen (ml. muunlaisten aineistojen kuvailutietojen) käsittely siten, että seuraavat periaatteet toteutuvat:

1. *Niin avointa kuin mahdollista, niin suljettua kuin välttämätöntä –periaate*
 - kaikki mikä voidaan tehdä julkiseksi ja uudelleen käytettäväksi myös tehdään
 - sellainen mitä ei voida julkistaa ja jakaa säilytetään turvallisesti

2. *FAIR-laatutavoitteet*
 - löydettävyys
 - saavutettavuus
 - yhteentoimivuus
 - uudelleenkäytettävyys

3. *Vastuullisuus*
 - tutkimusalan periaatteet
 - tutkimuseettiset periaatteet
 - lainsäädäntö
 - tietoturva ja tietosuojat

Tutkimusdatan saatavuuden osalinjaus: periaatteet

- *Periaate 1:* Kaikki keskeinen tutkimusdata avataan saataville siinä määrin ja niin pian kuin se vastuullisen datanhallinnan vaatimusten rajoissa on mahdollista. ("Datan avoimuuden periaate")
- *Periaate 2:* Tutkimustyössä noudatetaan hyviä tutkimusdatan hallinnan käytäntöjä. ("Hyvien käytäntöjen periaate")
- *Periaate 3:* Tutkijan arvioinnissa huomioidaan tutkimusdataan liittyvä työ, tutkijan ansiot hyvän datanhallinnan edistämisessä sekä soveltuvin osin tutkimusdatan julkistaminen. ("Arvostuksen periaate")
- *Periaate 4:* Tutkijoille tarjotaan asianmukainen hyvän datanhallinnan infrastruktuuri ja palvelut. ("Hyvien palvelujen periaate")
- *Periaate 5:* Tutkijoita tuetaan ja kannustetaan tutkimusdatan mahdollisimman suureen avoimuuteen sekä oman tutkimusalansa avoimuuden kulttuurin kehittämiseen. ("Avoimuuden kulttuurin periaate")

Tutkimusdatan saatavuuden osalinjaus: tavoitteet

- *Tavoite 1:* Tutkimusorganisaatiot sitoutuvat tukemaan tutkijoita hyvän tutkimusdatan periaatteiden noudattamisessa ja kannustavat tutkijoita jakamaan tietoaaineistojaan.
- *Tavoite 2:* Tutkimusyhteisössä on tarjolla kattavasti eri tieteenalojen ja tutkimuksen elinkaaren tarpeet huomioivaa tutkimusdatan hallinnan koulutusta, tukea ja osaamisen kehittämistä.
- *Tavoite 3:* Tutkimusorganisaatiot takaavat tutkijoilleen kaikelle Suomessa tuotetulle tai hyödynnettävälle tutkimusdatalle hyvän datanhallinnan periaatteiden kannalta asianmukaisen tallennuspaikan.
- *Tavoite 4:* Tutkimusdatan hallintasuunnitelmien laatiminen ja ylläpitäminen sekä tutkimusaineistojen dokumentointi ovat osa normaalia tutkimus- ja kehittämisprojektin laadunhallintaa. Aineistot dokumentoidaan siten, että rikas metadata tukee aineiston käyttöä ja löydettävyyttä.
- *Tavoite 5:* Tutkimusdatan oikeuksien hallinnan vastuut ja oikeudet on organisaatiossa ymmärretty ja sovittu. Tutkimusdatan käyttöön liittyvät oikeudet ja käytön ehdot sovitaan selkeästi kaikkien asianomaisten toimijoiden kesken.
- *Tavoite 6:* Tutkimusdataa arvostetaan ja arvioidaan tutkimusorganisaatioissa itsenäisenä tutkimustuotoksena.
- *Tavoite 7:* Kehitetään ja otetaan käyttöön tutkimusdatan avoimuuden kypsyystasomalli. Kaikki tutkimusorganisaatiot saavuttavat tässä arvioinnissa vähintään kohtuullisen tason.

Strategiset periaatteet

Periaate 1

Kaikki keskeinen tutkimusdata avataan saataville siinä määrin ja niin pian kuin se vastuullisen datanhallinnan vaatimusten rajoissa on mahdollista.

Selvennys: Tutkijan keräämä tai tuottama tietoaaineisto on lähtökohtaisesti niin avointa kuin mahdollista ja niin suljettua kuin välttämätöntä. Kaikkien aineistojen metadata on lähtökohtaisesti avointa, mikäli tälle ei ole erityistä estettä.

Periaate 2

Tutkimustyössä noudatetaan hyviä tutkimusdatan hallinnan käytäntöjä.

Selvennys: Tutkijat noudattavat alansa hyviä datanhallinnan käytäntöjä koko tutkimuksen elinkaaren läpi, mahdollistaen näin tutkimusdatan avoimen saatavuuden. Tutkijat ja tutkimuksen tukihenkilöt tuntevat FAIR-periaatteet ja osaavat soveltaa niitä. Tutkijat noudattavat erityistä tarkkuutta käsitellessään sensitiivistä tietoa käsittelevää tutkimusdataa. Tukena voi käyttää täydentävää ohjetta arkaluonteisia ja luottamuksellisia tietoja sisältävän datan hallinnan suunnitteluun.

Periaate 3

Tutkijan arvioinnissa huomioidaan tutkimusdataan liittyvä työ, tutkijan ansiot hyvän datanhallinnan edistämisessä sekä soveltuvin osin tutkimusdatan julkistaminen.

Selvennys: Tutkimustyössä osoitettu hyvien datanhallinnan käytäntöjen noudattaminen ja kehittäminen sekä avoimen saatavuuden edistäminen hyödyttää tutkijaa hänen urallaan. Tutkijan arviointia kehitetään hyödyntämällä kansallista suositusta vastuulliseen tutkijanarviointiin. Meritoitumisen ja kannustimien kehittäminen toteutetaan niin, että sensitiivisen tai luottamuksellisen aineiston käyttö ei toisaalta ole epäedullista tutkijan uran kannalta.

Periaate 4

Tutkijoille tarjotaan asianmukainen hyvän datanhallinnan infrastruktuuri ja palvelut.

Selvennys: Tutkijoiden saatavilla on riittävät palvelut datan avoimen saatavuuden mahdollistamiseksi. Tutkimusdataan liittyviä tukipalveluita kehitetään moniammatillisessa yhteistyössä. Kehittämisessä huomioidaan avoimuuden myötä syntyvät uudet tarpeet ja kansainväliset käytännöt yhteensopivuuden ja löydettävyyden takaamiseksi.

Periaate 5

Tutkijoita tuetaan ja kannustetaan tutkimusdatan mahdollisimman suureen avoimuuteen sekä oman tutkimusalansa avoimuuden kulttuurin kehittämiseen.

Selvennys: Tutkimusdatan avoimen saatavuuden edistäminen on koko tutkimusyhteisön yhteinen tavoite. Tutkimusorganisaatiot sitoutuvat tähän tavoitteeseen, ja tutkijoille tarjotaan koulutusta ja tukea organisaatiossa sekä mahdollisesti tieteenalalla noudatettavien käytäntöjen osalta. Tutkijoita kannustetaan myös osallistumaan aktiivisesti oman tieteenalansa standardien, sanastojen ja muiden datan avoimuuden työkalujen kehittämiseen.

Tavoitteet 2021-2025

Tavoite 1

Tutkimusorganisaatiot sitoutuvat tukemaan tutkijoita hyvän tutkimusdatan hallinnan periaatteiden noudattamisessa ja kannustavat tutkijoita jakamaan tietoaaineistojaan.

Tavoitteen vaatimat toimenpiteet:

- Tutkimusorganisaation johto sitoutuu edistämään hyvän tutkimusdatan hallinnan vaatimaa kulttuurinmuutosta. Vuonna 2021 tutkimusorganisaatiot nimeävät tehtävälle vastuutahon, joka kuuluu organisaation johtoryhmään (tai vastaavaan) tai raportoi suoraan tälle.
- Vuoteen 2022 mennessä tutkimusorganisaatiot laativat tiekartan kohti organisaation yhteisiä hyvän tutkimusdatan käytäntöjä. Tiekarttaa laadittaessa on arvioitava myös tulevien muutosten vaatimat resurssit.
- Vuoteen 2023 mennessä tutkimusorganisaatiot laativat kansallisen tutkimusaineistojen avoimuuden kypsyystasomallin pohjalta omat hyvän datanhallinnan laatuindikaattorinsa, joita soveltamalla seuraavat organisaationsa edistymistä hyvän tutkimusdatan hallinnan alueella.

Tavoite 2

Tutkimusyhteisössä on tarjolla kattavasti eri tieteenalojen ja tutkimuksen elinkaaren tarpeet huomioivaa tutkimusdatan hallinnan koulutusta, tukea ja osaamisen kehittämistä.

Tavoitteen vaatimat toimenpiteet:

- Organisaatiot selvittävät tarpeet datanhallinnan koulutukselle vuoteen 2022 mennessä.
- Organisaatiot tarjoavat itsenäisesti tai yhteistyössä yhteisönsä tarpeita vastaavasti hyvän datanhallinnan koulutusta kaikkien uravaiheiden tutkijoille ja opiskelijoille, opettajille, ohjaajille sekä tukihenkilöstölle vuoteen 2023 mennessä.
- Organisaatiot kehittävät ja tukevat uudenlaisia palveluita ja ammatillisia rooleja, jotka mahdollistavat vastuullisen datanhallinnan sekä avoimen ja FAIR-datan tuottamisen.
- Datanhallinnan koulutus ja osaamisen kehittäminen toteutetaan moniammatillisessa yhteistyössä ja siinä otetaan huomioon tutkimusalojen erityispiirteet, FAIR-periaatteet ja tietoaaineiston koko elinkaari.
- Osaamisen ja palveluiden rakenteita ja sisältöjä kehitetään kansallisessa ja kansainvälisessä yhteistyössä.

Tavoite 3

Tutkimusorganisaatiot takaavat tutkijoilleen kaikelle Suomessa tuotetulle tai hyödynnettävälle tutkimusdatalle hyvän datanhallinnan periaatteiden kannalta asianmukaisen tallennuspaikan.

Tavoitteen vaatimat toimenpiteet:

- Avoimen tieteen koordinaatio käynnistää vuoden 2021 aikana suomalaisten tutkimusorganisaatioiden yhteisen selvityshankkeen, jonka tehtävänä on laatia suositus tutkimusdatan tutkimusprojektin aikaiseen tallentamiseen ja käsittelyyn käytettävistä kansallisista tai kansainvälisistä palveluista. Selvityksen määräaika on vuosi 2022.
- Selvitykseen liittyen suomalaiset tutkimusorganisaatiot analysoivat tuottamansa ja käyttämänsä nykyiset tallennuspalvelut sekä niiden tukipalvelut FAIR-periaatteiden ja datapolitiikkojen toteuttamisen näkökulmasta sekä kartoittavat olemassa oleviin palveluihin liittyvät puutteet. Havaittujen puutteiden osalta organisaatiot käynnistävät tarvittavat kehittämistoimenpiteet.
- Em. selvitysten ja suositusten pohjalta tutkimusorganisaatiot tuottavat ja tarjoavat vuoteen 2024 mennessä tutkijoille ja tutkimusryhmille ohjeistuksen sekä riittävät tukipalvelut hyvän datanhallinnan periaatteiden mukaiseen tiedontallennukseen.
- Tietoturvallisten ja FAIR-periaatteiden mukaisten tallennuspalveluiden kehittämisessä ja toteuttamisessa tutkimusorganisaatiot huomioivat käyttäjälähtöisyyden keskeisenä periaatteena.

Tavoite 4

Tutkimusdatan hallintasuunnitelmien laatiminen ja ylläpitäminen sekä tutkimusaineistojen dokumentointi ovat osa normaalia tutkimus- ja kehittämisprojektin laadunhallintaa. Aineistot dokumentoidaan siten, että rikas metadata tukee aineiston käyttöä ja löydettävyyttä.

Tavoitteen vaatimat toimenpiteet:

Vuoteen 2023 mennessä tutkimusorganisaatiot:

- sisällyttävät datanhallintasuunnitelman laadinnan ja ylläpidon osaksi tutkimus- ja palveluprosesseja aineiston elinkaaren jokaisessa vaiheessa
- tarjoavat ohjeet, käytänteet ja koulutusta aineistohallinnan suunnitteluun opiskelijoille, tutkijoille ja muulle henkilöstölle
- huolehtivat, että opinnäytetöiden ohjaajat osaavat arvioida ja kommentoida datanhallintasuunnitelmia
- tarjoavat tutkimusaineistojen dokumentoinnin koulutusta sekä työkaluja opiskelijoille ja tutkijoille
- tukevat tutkimuksen suunnittelua siten, että myös aineistojen dokumentoinnin kustannuksiin osataan varautua.

Tavoite 5

Tutkimusdatan oikeuksien hallinnan vastuut ja oikeudet on organisaatiossa ymmärretty ja sovittu. Tutkimusdatan käyttöön liittyvät oikeudet ja käytön ehdot sovitaan selkeästi kaikkien asianomaisten toimijoiden kesken.

Tavoitteen vaatimat toimenpiteet:

- Avoimen tieteen koordinaatio käynnistää vuoden 2021 aikana tutkimusdatan immateriaalioikeuksiin liittyvän selvityksen, jonka tarkoituksena on selkeyttää omistajuus- ja sopimuskysymyksiä. Selvityksen tuloksena laaditaan kansallisia suosituksia sopimusmalleiksi.
- Vuoteen 2022 mennessä tutkimusrahoittajat ja -organisaatiot ohjeistavat tutkijoita, että digitaaliset tietoaineistot tulee joko lisensoida tarkoituksenmukaisella avoimuuden periaatteet toteuttavalla ihmis- ja koneluettavalla lisenssillä tai tutkimusdatan yhteiskunnallinen vaikuttavuus tulee saavuttaa kaupallistamisella.
- Tutkimusorganisaatioissa tarjotaan vuoteen 2023 mennessä kattavasti moniammatillisessa yhteistyössä opetusta, koulutusta ja neuvontaa tutkimusdataan kohdistuvista oikeuksista ja avoimen tieteen lisensseistä.
- Tutkimusaineistojen oikeuksista ja vastuista sovitaan vuoteen 2023 mennessä kaikissa tutkimusorganisaatioissa selkeästi.

Tavoite 6

Tutkimusdataa arvostetaan ja arvioidaan tutkimusorganisaatioissa itsenäisenä tutkimustuotoksena.

Tavoitteen vaatimat toimenpiteet:

- Vuoteen 2022 mennessä tutkimusorganisaatiot:
 - laativat ohjeistuksen hyvän datanhallinnan alueen ansioiden huomioon ottamisesta rekrytointi-, meritoitumis- ja palkitsemiskäytännöissään.
- Vuoteen 2024 mennessä tutkimusorganisaatiot:
 - kartoittavat perustiedot tietoaineistoistaan ja laativat dokumentoidut prosessit näiden tietojen ylläpitämiseksi
 - tarjoavat julkaistaville/arkistoitaville tietoaineistoilleen tarvittavat kuratointipalvelut, so. palvelut aineistojen viimekätiseen laadunvarmistukseen ennen julkaisua/arkistointia (pysyväistallennusta)
 - laativat organisaatiossa noudatettavan, kansallisten ja kansainvälisten käytäntöjen mukaisen ohjeistuksen tietoaineistoihin viittaamiselle.

Tavoite 7

Kehitetään ja otetaan käyttöön tutkimusdatan avoimuuden kypsyystasomalli. Kaikki tutkimusorganisaatiot saavuttavat tässä arvioinnissa vähintään kohtuullisen tason.

Tavoitteen vaatimat toimenpiteet:

- Vuosi 2020: Kypsyystasomallia valmistelevan työryhmän asettaminen.
- Vuosi 2021: Työryhmän esitys ja sen vahvistaminen.
- Vuosi 2022: Mallin käyttöönotto.
- Vuosi 2025: Kaikki tutkimusorganisaatiot saavuttavat asetetun tavoitetason.

Toimenpiteiden 2021-2025 aikajana

2021:

- Organisaatioiden tutkimusdatan avoimuuden vastuutahojen nimeäminen
- Yhteinen esitys tutkimusdatan hallinnan kypsyystasomallista ja sen vahvistaminen
- Yhteinen selvitys tutkimusdatan immateriaalioikeuksista ja suositukset sopimusmalleiksi

2022:

- Yhteisen kypsyystasomallin käyttöönotto
- Yhteinen selvitys tutkimusdatan tallentamisen palveluista
- Organisaatioiden hyvän tutkimusdatan hallinnan tiekartat
- Organisaatioiden selvitykset datanhallinnan koulutustarpeista
- Organisaatioiden ohjeistukset datanhallinnan alueen ansioiden huomioon ottamisesta

2023:

- Organisaatioiden hyvän tutkimusdatan hallinnan laatuindikaattorit
- Organisaatioiden datanhallinnan koulutukset suunniteltu ja toteutuksessa
- Organisaatioiden aineistojen dokumentoinnin ja datanhallintasuunnitelmien käytännöt
- Organisaatioiden käytännöt tutkimuksen tietoaineistojen oikeuksista ja vastuista

2024:

- Organisaatioiden tutkimusdatan tiedontallennuksen ohjeistukset ja tukipalvelut toteutettu
- Organisaatioiden tutkimuksen tietoaineistot kartoitettu ja kartoituksen ylläpito suunniteltu

2025:

- Organisaatiot saavuttavat tutkimusdatan hallinnassa vähintään kohtuullisen kypsyystason

Linjauksen toteutus

- Linjauksen toteutusta koordinoi Avoimen Tieteen kansallisen koordinaation yhteyteen perustettava toteutusryhmä, jolle osoitetaan asianmukaiset resurssit yhteydenpitoon tutkimusorganisaatioihin, kansallisen tilanteen kartoittamiseen, tarvittavien yhteisten selvitysten ja suositusten laatimiseen sekä toimenpiteiden toteutumisen seuraamiseen.

Kiitos!

Kysymyksiä?