

HUOM! TÄMÄ ON RINNAKKAISTALLENNE

Rinnakkaistallennettu versio *voi* erota alkuperäisestä julkaistusta sivunumeroiltaan ja ilmeeltään.

Tekijä(t): Väänänen, Kristiina; Muhonen, Tiina; Varis, Kaisa

Otsikko: Tutkimusaineiston avaaminen – vinkit tutkijoille

Versio: Kustantajan julkaisema artikkeli

Käytä viittauksessa alkuperäistä lähdettä:

Väänänen, K., Muhonen, T., Varis, K. (2024). Tutkimusaineiston avaaminen – vinkit tutkijoille Pulssi-portaali 16.9.2024. Karelia-ammattikorkeakoulu
<https://www.karelia.fi/2024/09/tutkimusaineiston-avaaminen-vinkit-tutkijoille/>

PLEASE NOTE! THIS IS PARALLEL PUBLISHED VERSION / SELF-ARCHIVED VERSION OF THE OF THE ORIGINAL ARTICLE

This is an electronic reprint of the original article.

This version *may* differ from the original in pagination and typographic detail.

Author(s): Väänänen, Kristiina; Muhonen, Tiina; Varis, Kaisa

Title: Tutkimusaineiston avaaminen – vinkit tutkijoille

Version: Publisher's pdf

Please cite the original version:

Väänänen, K., Muhonen, T., Varis, K. (2024). Tutkimusaineiston avaaminen – vinkit tutkijoille Pulssi-portaali 16.9.2024. Karelia-ammattikorkeakoulu
<https://www.karelia.fi/2024/09/tutkimusaineiston-avaaminen-vinkit-tutkijoille/>

Tutkimusaineiston avaaminen – vinkit tutkijoille

16.9.2024

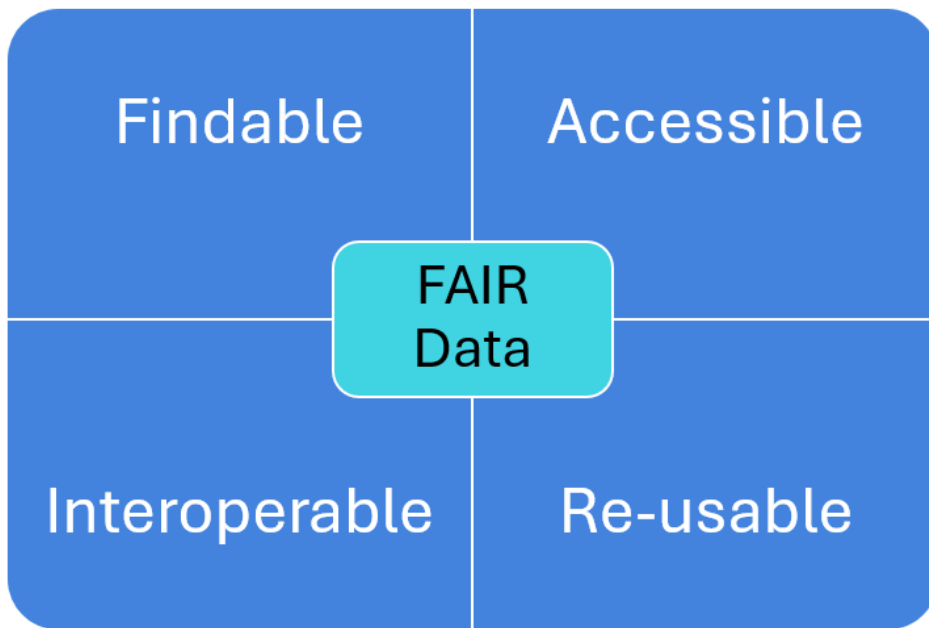
Avoin TKI-toiminta tarkoittaa avoimien toimintamallien hyödyntämistä ammattikorkeakoulujen tutkimus- kehittämis- ja innovaatiotoiminnassa. Avoimuuden muotoja ovat mm. tutkimustulosten avoin julkaiseminen, tutkimusaineistojen avaaminen sekä avoimet tutkimusmenetelmät.

Tutkimusaineistojen, kuten kysely- ja haastatteluaineistojen tai mittaus- ja havainnointidatan avoimuus lisää tutkimuksen laatua, läpinäkyvyyttä ja luotettavuutta. Aineiston avaaminen mahdollistaa myös sen jatkokäytön uusissa tutkimuksissa. Tutkimusaineiston avaaminen vaatii tarkkaa suunnittelua jo hyvissä ajoin ennen sen keräämistä. Huomioon on otettava mm. tutkimukseen osallistuvien henkilöiden suostumus, lailliset vaatimukset kuten EU-alueen tietosuojalainsäädäntö (GDPR) sekä rahoittajien asettamat ehdot.

[INVEST4EXCELLENCE-projektissa](#) yhtenä tavoitteena on pilotoida käytänteitä tutkimuksen avoimuuden lisäämiseksi. Avoimen julkaisutoiminnan lisäksi pääsimme hankkeessa selvittämään ja kokeilemaan, mitä tutkimusaineiston avaaminen on käytännössä. Mitä seikkoja pitää ottaa huomioon jo tutkimusta suunniteltaessa ja miten aineiston avaaminen käytännössä toteutetaan. Tämä artikkeli antaa vinkkejä muille tutkijoille siitä, miten ja millaisia aineistoja tutkimusprojekteista voi avata.

Avoimuuden FAIR-periaatteet

Kaikessa tutkimuksessa on huomioitava, että kerätyt tietoaineistot ovat saavutettavia. Tämä tarkoittaa, että ihmisten ja järjestelmien tulee löytää aineistot ja niissä oleva tieto sekä osata analysoida ja yhdistellä sitä. Tätä tavoitetta kuvataan usein [FAIR-periaatteiden](#) avulla (Findable, Accessible, Interoperable, Re-usable). FAIR-periaatteita toteuttaen aineistot ovat turvassa ja niillä on pysyvät tunnisteet, jolloin aineistoon voidaan viitata muissa tutkimuksissa. Huomioithan FAIR-periaatteet jo tutkimusta suunnitellessasi, jolloin aineiston jatkokäsittely myöhemmin on helpompaa ja aineiston löydettävyyttä sekä käytettävyyttä paranevat.



Kuva 1. Tutkimusaineiston FAIR-periaatteet (löydettävä, saavutettava, yhteentoimiva ja uudelleenkäytettävä).

Mikäli suunnittelet tutkimusaineiston avaamista ensimmäistä kertaa, muistathan, että tutkimusorganisaatioista löytyy asiantuntijoita ja tukipalveluita, jotka voivat auttaa tutkimuksen eri vaiheissa.

1. Tutkimuksen suunnittelu ja toteutus

Karelia-ammattikorkeakoulu edellyttää datapolitiikassaan kaikilta tutkimus- ja kehittämishankkeilta aineistohallintasuunnitelman laatimista heti hankkeen aloitusvaiheessa. [Aineistohallintasuunnitelman](#) kirjataan mm. mitä aineistoja hankkeessa käsitellään, miten aineistot säilytetään ja mitä sille tehdään hankkeen päätyttyä. Suunnitelma kattaa aineiston koko elinkaaren aineiston keruusta ja käsittelystä sen julkaisemiseen tai arkistointiin jatkokäyttöä varten.

Useat suomalaiset korkeakoulut – kuten myös Karelia – suosittelevat [DMPTuuli](#) -järjestelmän käyttöä aineistohallintasuunnitelman luomisessa ja osa rahoittajista vaatii suunnitelmaa jo projektin hakuvaiheessa. Tutkimusta suunniteltaessa on myös jo hyvä pohtia, onko aineistossa sellaista arkaluontoista tietoa, ettei sitä kannata avata tai onko aineiston avaamiselle muita eettisiä, laillisia tai sopimusoikeudellisia esteitä. Tällaisissa tapauksissa voidaan myös julkaista pelkästään aineiston kuvailutieto (metadata).

Tutkimuksen suunnittelu- ja toteutusvaiheessa on huomioitava se, että vain anonymia dataa voi avata ja datan säilytyksestä ja mahdollisesta avaamisesta on informoitava tutkittavia. Jotta tutkittavien suostumus myös datan jatkokäyttöön ja avaamiseen voidaan varmistaa, tutkittavien on oltava rekrytointivaiheessa tietoisia siitä, että tutkimuksessa kerätty aineisto voidaan anonymisti avata. Tutkittavia henkilöitä voidaan informoida datan jatkokäytöstä ja

avaamisesta esimerkiksi tutkimuslupakirjeellä, joka toimitetaan heille hyväksyttäväksi rekrytointivaiheessa.

Yhä useammin myös rahoittaja voi vaatia aineiston avaamista ainakin osittain, mikäli se on mahdollista. Viimeistään tutkimusta suunniteltaessa on hyvä pohtia myös se, miten aineistoa säilytetään ja käsitellään turvallisesti.

2. Aineiston käsittely

Peruseriaatteena on, että avattava aineisto on alkuperäinen aineisto, jota muut tutkijat voivat hyödyntää. Mikäli aineistoa on käsitelty, esimerkiksi käännetty toiselle kielelle, tulee tämä ilmaista selkeästi aineiston metatiedoissa. Tallenna aineisto sellaisessa muodossa, että se on helposti muiden käytettävissä. INVEST4EXCELLENCE-tutkimuksemme oli laadullinen haastattelututkimus, jossa litteroitu aineisto käännettiin englanniksi, anonymisoitiin ja jäsenneltiin temaattisesti taulukkolaskentaohjelman avulla.

Anonyymiuden varmistaminen on tutkimusaineiston avaamisen haastavin osa, koska anonymisointiprosessi on suunniteltava aineiston piirteiden perusteella. Sovelsimme tutkimuksemme anonymisoinnin suunnittelussa [Tietoarkiston aineistohallinnan käsikirjan ohjeita tunnisteellisuudesta ja anonymisoinnista](#). Olennaista tässä prosessissa on huomioida, voidaanko henkilö tunnistaa suoraan esimerkiksi henkilötietojen avulla tai epäsuorasti yhdistelemällä erilaisia aineistosta tai muista lähteistä löytyviä tietoja.

INVEST4EXCELLENCE-tutkimusaineistomme anonymisointiprosessissa [suorat henkilötiedot](#), kuten nimi, työpaikka, asema tai ammattinimike, oli yksikertaista poistaa aineistosta. Kuitenkin pienessä, tiettyyn ammattiryhmään kohdistuneessa ja alueellisesti rajautuneessa otannassa henkilö voi olla [tunnistettavissa muutenkin](#) kuin esimerkiksi nimen tai yrityksen perusteella. Tutkimuksen kohteena oleva henkilö voi esimerkiksi kertoa eri kohdissa aineistoa a) toimivansa tietyllä, alueellisesti pienellä toimialalla, b) toimivansa Ilomantsin alueella ja c) tekevänsä tiivistä yhteistyötä tietyn yrityksen kanssa. Tässä tapauksessa henkilö voi olla tunnistettava tietylle ihmisryhmälle, koska esimerkiksi toimiala ja tutkimuksen alueellinen konteksti, ja niiden yhdistelmät vaikuttavat siihen, onko tutkimukseen osallistunut henkilö tunnistettava aluetta ja sen toimialarakennetta tuntevalle henkilöryhmälle. Tämän takia anonymisoinnissa on tarkasteltava aineistokohtaisesti, miten erilaiset taustatiedot tulevat aineistossa kokonaisuutena esiin ja vaikuttavat tunnistettavuuteen.

INVEST4EXCELLENCE-tutkimuksessamme anonymisointiprosessi oli kaksivaiheinen. Ensimmäisenä ensisijainen tutkija suunnitteli ja toteutti anonymisoinnin. Tämän jälkeen toinen toimialaa tunteva tutkija, joka ei osallistunut haastatteluaineiston keräämiseen, tarkisti aineiston pohtien koko ajan sitä, onko siellä vielä mahdollisia tunnistetietoja.

Yleisten anonymisointiohjeiden lisäksi kansainvälisillä avoimilla data-arkistoilla voi olla olemassa erityisiä anonymisointia koskevia ohjeita, jotka on hyvä ottaa huomioon anonymisoinnin suunnitteluprosessissa.

3. Data-arkiston valitseminen ja aineiston lisensointi

Tutkimusaineistot avaamiseen on olemassa lukuisia kansallisia ja kansainvälisiä data-arkistoja. Sopivan data-arkiston valinnassa on hyvä huomioida etenkin seuraavat asiat:

- arkisto on (oman alan) tutkijoiden tuntema ja käyttämä,
- arkisto antaa aineistolle pysyvän tunnisteeseen, kuten DOI tai URN tai signum,
- arkistolla on luotettavuudesta kertova sertifikaatti,
- arkisto antaa valita käyttöehdot (lisenssin), joilla aineisto on jatkokäytettävissä.

Yleisimpiä data-arkistoja on esitelty Karelian avoimen TKI-toiminnan [oppaassa](#). Arkistoja voi etsiä myös [re3data-rekisteristä](#). Karelian oma pitkäaikaissäilytykseen tarkoitettu verkkolevy soveltuu sellaisten aineistojen säilytykseen, joita ei julkaista avoimesti, vaan ne ovat käytettävissä esimerkiksi erillisellä luvalla. Tutkimusaineistosta kannattaa kuitenkin tehdä julkinen kuvailutieto myös silloin kun aineistoa ei voi avata.

Tässä vaiheessa määritellään myös se, miten jakamaasi aineistoa saa jatkokäsitellä ja millaisilla ehdoilla sitä voi hyödyntää. Yleisesti käyttöoikeudet määritellään [Creative Commons 4.0](#) -lisenssien avulla (CC-lisenssit). Avoimen tieteen ja tutkimuksen ohjeissa suositeltu jakamistapa on [CC-BY-4.0 -lisenssi](#), jolloin aineistoa voi muokata (muokkauksesta mainiten) ja jakaa vapaasti, mutta alkuperäinen lähteet on aina mainittava. Tämä lisenssi sallii myös kaupallisen hyödyntämisen. Korkeakoulujen tutkijoiden kannattaa huomioida, että myös jatkuvan oppimisen palvelut ovat kaupallista hyödyntämistä, jolloin kaupallisen hyödyntämisen mahdollistaminen voi usein olla tarkoituksenmukaista.

INVEST4EXCELLENCE:ssä päädyimme viiden eurooppalaisen korkeakoulun kesken avaaman tutkimusaineistot [Zenodo](#)-arkistossa, mikä on kansainvälinen arkisto EU-rahoitteisten projektien tuotoksille.

4. Aineistojen julkaiseminen ja aineistosta viestiminen

Kun aineistolle on luotu tarvittava metadata ja lisenssi, se on valmis julkaistavaksi data-arkistossa. INVEST4EXCELLENCE-tutkimuksessa käyttämämme data-arkisto Zenodo antaa avatulle datalle pysyvän tunnisteeseen (DOI), joka helpottaa sen löytämistä ja siihen viittaamista. Pysyvän tunnisteeseen kautta saimme datan jaettua tutkimusjulkaisuumme tausta-aineistona. Lisäksi pysyvä tunniste mahdollisti suoran viittaamisen tutkimuksessa esiin nostettuihin esimerkkeihin.

Kansainvälisesti avattu data julkaistiin metatietoineen myös Karelian Repotronic-järjestelmässä, joiden kautta se on näkyvillä Karelian kotisivuilla sekä [Tiede ja tutkimus-portaalissa](#).

Lisätietoja tutkimusaineiston avaamisesta voi lukea [Oppaasta avoimeen TKI-toimintaan Karelia-ammattikorkeakoulussa](#).

INVEST4EXCELLENCE-projekti on [INVEST-yliopistoallianssin](#) tutkimuspainotteinen liitännäisprojekti, jonka rahoitus tulee Horisontti2020-rahoitusohjelmasta. Karelia-

ammattikorkeakoulun vastuulla on ollut erityisesti tuottaa strategista ohjausta TKI-toiminnan yhteiseen suunnitteluun sekä kehittää INVEST:in toimintaa tukevia työkaluja henkilöstön osaamisen vahvistamiseen. INVEST-yliopistoallianssi on seitsemän eurooppalaisen korkeakoulun verkosto, jonka toimintaa ohjaa yhteinen visio kohti kestävämpää ja vastuullisempaa Eurooppaa.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101035815

Kirjoittajat:

Kristiina Väänänen, projektipäällikkö, Karelia-ammattikorkeakoulu
Tiina Muhonen, projektiasiantuntija, Karelia-ammattikorkeakoulu
Kaisa Varis, tietoasiantuntija, Karelia-ammattikorkeakoulu

Lähteet:

Creative Commons. CC BY 4.0 Attribution 4.0 international.
<https://creativecommons.org/licenses/by/4.0/deed.en>. Viitattu 11.9.2024.

Creative Commons Suomi. <https://creativecommons.fi/>. Viitattu 11.9.2024.

DMPTuuli. Aineistonhallintasuunnitelma. <https://www.dmptuuli.fi/>. Viitattu 11.9.2024.

Fairdata.fi. FAIR-periaatteet. <https://www.fairdata.fi/tietoa-fairdatasta/fair-periaatteet/>. Viitattu 15.8.2024.

INVEST. Innovations of Regional Sustainability: European University Alliance. <https://www.karelia.fi/invest/>. Viitattu 28.8.2024.

INVEST4EXCELLENCE (2024). Horisontti2020 SwafS -projekti 101035815. <https://www.invest4excellence.eu/>. Viitattu 15.8.2024.

Karelia-ammattikorkeakoulu. Opas avoimeen TKI-toimintaan Karelia-ammattikorkeakoulussa. <https://libguides.karelia.fi/c.php?g=670780&p=4762766>. Viitattu 11.9.2024.

Re3data.org. Registry of research data repositories. <https://www.re3data.org/>. Viitattu 28.8.2024.

Suomen Akatemia. Aineistonhallintasuunnitelma: Aineiston hallinnan suunnittelu. <https://www.aka.fi/tutkimusrahoitus/hae-rahoitusta/nain-haet-rahoitusta/ohjehakemisto/aineistonhallinta/aineistonhallintasuunnitelma/>. Viitattu 15.8.2024.

Tiedejatutkimus.fi. Hae tietoa tutkimuksesta Suomessa. <https://tiedejatutkimus.fi/fi/>. Viitattu 11.9.2024.

Tietoarkisto. Tunnisteellisuus ja anonymisointi. <https://www.fsd.tuni.fi/fi/palvelut/aineistonhallinta/tunnisteellisuus-ja-anonymisointi/>. Viitattu 11.9.2024.

Tietosuojavaltuutetun toimisto. Mikä on henkilötieto? <https://tietosuoja.fi/mika-on-henkilotieto>. Viitattu 28.8.2024.

Tietosuojavaltuutetun toimisto. Pseudonymisoidut ja anonymisoidut tiedot. <https://tietosuoja.fi/pseudonymisointi-anonymisointi>. Viitattu 28.8.2024.

Zenodo. <https://zenodo.org/>. Viitattu 11.9.2024.