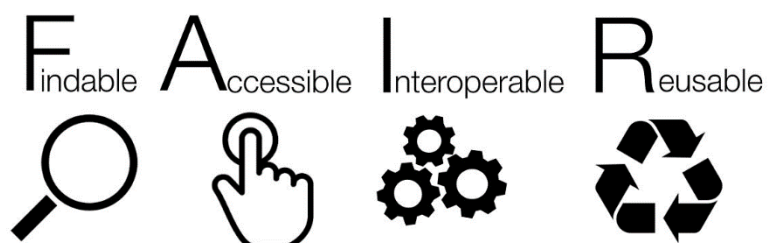




Avoin tiede

Interoperable₁₋₂₋₃

Ohjeita tutkimusdatan yhteentoimivuuden edistämiseen





Sisälllys

Johdanto.....	1
Tutkijalle semanttisen yhteentoimivuuden edistämisestä	1
Datanhallinnan asiantuntijalle semanttisen yhteentoimivuuden edistämisestä.....	2
Palveluntarjoajalle semanttisen yhteentoimivuuden edistämisestä	2
Lisämateriaalia	2
11 Aineistot ja metatiedot ovat sisällöltään määrämuotoisia, monikäyttöisiä, saatavilla olevia ja jaettua kieltä käyttäviä.....	3
Ohjeita soveltamiseen tutkijalle.....	3
Ohjeita soveltamiseen datanhallinnan asiantuntijalle	3
Ohjeita soveltamiseen palveluiden tuottajalle	4
12. Aineistoissa ja metatiedoissa käytetään sanastoja, jotka noudattavat FAIR-periaatteita.....	4
Ohjeita soveltamiseen tutkijalle.....	4
Ohjeita soveltamiseen datanhallinnan asiantuntijalle	5
Ohjeita soveltamiseen palveluiden tuottajalle	5
13. Aineistoissa ja metatiedoissa on tyyppiteltyjä viittauksia muihin resursseihin	6
Ohjeita soveltamiseen tutkijalle.....	6
Ohjeita soveltamiseen datanhallinnan asiantuntijalle	6
Ohjeita soveltamiseen palveluiden tuottajalle	7
Laatijat.....	8

Johdanto

Yhteentoimivuus on yksi neljästä [FAIR-periaatteesta](#), joilla pyritään edistämään digitaalisten tutkimusaineistojen löydettävyyttä, saatavuutta, yhteentoimivuutta ja uudelleen käytettävyyttä. Ne muodostavat tavoitteet, joihin usein viitataan tutkimusdatan hallinnasta puhuttaessa. On kuitenkin tärkeä muistaa, että ne ovat teknisen avoimuuden väline, jota voidaan soveltaa eri tavoin erilaisissa yhteyksissä. FAIR-periaatteet ovatkin lähinnä työkalu, jonka avulla tutkimusaineistojen laatua voidaan parantaa, mutta eri tilanteissa soveltaminen voi tapahtua eri tavoin. Teknisesti täysin avoin muoto, koneluettavuus ja yhteentoimivuus voivat olla kalliita ja vaatia joskus huomattavia panostuksia, jotka eivät aina ole tarkoituksenmukaisia. Joissain tilanteissa on syytä tyytyä periaatteiden osittaiseen noudattamiseen. FAIR -periaatteet ovat kuitenkin aina hyvä ohjenuora, jos haluaa edistää aineiston löydettävyyttä ja käytettävyyttä.

FAIR-periaatteiden *Interoperable* -kohta sisältää kolme eri periaatetta, joiden soveltamista käsitellään tässä ohjeessa. Fokus on lähinnä semanttisessa yhteentoimivuudessa, joka on yhteentoimivuuden neljästä tasosta ehkä monimutkaisin tutkimusaineistojen teknisen avoimuuden kannalta ja se taso, johon FAIR-periaatteet keskittyvät.

Voit kokeilla [FAIR-Aware -arviointityökalua](#) tutustuaksesi aiheeseen.

Tutkijalle semanttisen yhteentoimivuuden edistämisestä

Laadukkaasti avatut ja viittauksia keränneet tutkimusaineistot ovat tutkijoille meriitti. Tutkimusaineistojen näkyvyys lisää yhteistyömahdollisuuksia. Datan jakaminen voi lisätä yhteiskunnallista ja muuta vaikuttavuutta tutkijan tutkimushankkeille. Aineistojen uudelleen käyttö lisää niiden arvoa. Tutkijoiden tulisi voida tulkita, käsitellä ja tallentaa olemassa olevia tutkimusaineistoja ja yhdistää niitä toisiin aineistoihin. Tämä pitäisi olla mahdollista tehdä myös automatisoidusti ilman tapauskohtaista erillisen lisätiedon hankkimista aineiston tarkoituksesta, rakenteesta, käytetyistä termeistä tai tiedostomuodoista.

Semanttinen yhteentoimivuus tarkoittaa sitä, että tietoja pystytään käsittelemään eri järjestelmissä, siirtämään järjestelmästä toiseen ja tulkitsemaan niitä ilman, että tiedon merkitys muuttuu tai vääristyy. Semanttisesti yhteentoimivaa dataa voi esimerkiksi yhdistellä tai käsitellä valituilla tutkimusmenetelmillä ja -ohjelmistoilla, vaikka ne tulevat eri lähteistä. Tämä edellyttää hyvää



aineistohallintaa sekä yleisesti käytettyjen sanastojen ja muiden yhteentoimivuutta edistävien työkalujen ja menetelmien hyödyntämistä.

Tässä ohjeessa kerrotaan mitä eri toimijat voivat tehdä edistääkseen aineistojen yhteentoimivuutta ja samalla parantaa niiden löydettävyyttä ja käytettävyyttä. Keskeinen osa yhteentoimivuuden toteuttamisesta saavutetaan hyödyntämällä olemassa olevia palveluja ja parhaita käytäntöjä. Tutkimusprosessissa tulee olla mukana datan hallinnan ja tarvittaessa datatieteiden osaamista.

Datanhallinnan asiantuntijalle semanttisen yhteentoimivuuden edistämisestä

Datan yhteentoimivuutta voidaan edistää ennen kaikkea hyvällä aineistohallinnan suunnittelulla ja hyödyntämällä yhteisiä standardeja, käytäntöjä ja ratkaisuja. Datanhallinnan asiantuntijan ja kouluttajan tehtävä on tutkijoiden tukeminen oikeiden formaattien, palveluiden ja kontrolloitujen sanastojen löytämisessä. Tutkija tuntee oman alansa käytäntöjä ja datan laadun parantaminen FAIR-periaatteita soveltamalla lisää aineiston löydettävyyttä ja käytettävyyttä. Data-asiantuntija tukee tutkijaa vielä parempien datanhallinnan käytäntöjen äärelle ja huolehtii vuorovaikutuksesta palveluiden ja sanastojen kehittäjien suuntaan.

Data-asiantuntijan tulee ymmärtää metadatan, tunnisteiden ja linkkaarenhallinnan merkitys tutkimuksen toistettavuudelle sekä dataan viittaamisen käytännöt, ottaen huomioon viittaamisen mm. julkaisuista dataan, datasta dataan tai datasta julkaisuihin.

Palveluntarjoajalle semanttisen yhteentoimivuuden edistämisestä

Semanttisen yhteentoimivuuden kehittäminen vahvistaa palvelun vaikuttavuutta ja käyttäjäystävällisyyttä. Palveluilla on hyvin tärkeä rooli, ei vain FAIR -periaatteiden toteutumisen mahdollistajana, vaan myös niiden aktiivisena edistäjänä. Työtä on tehtävä jatkuvasti ja käyttäjälähtöisesti, aina huomioiden myös ulkoiset vaatimukset. Tarjoamalla pysyviä tunnisteita, avoimia rajapintoja ja integroitua referenssimetadatoja palvelu voi tukea tutkimusyhteisöä huomattavasti ja tavalla, joka ei edes ole mahdollista ilman tätä tukea. Palvelun tarjoaja voi hyödyntää jotain FAIR-evaluointityökalua parantaakseen tietojensa koneluettavuutta.

Lisämateriaalia

FAIR-periaatteet: <https://www.go-fair.org/fair-principles/>

FAIRsFAIR: Framework for assessing FAIR services <https://doi.org/10.5281/zenodo.5336234>



F-UJI tool: <https://www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool>

Pitkäaikaissäilytyksen vaatimukset: <https://digitalpreservation.fi/specifications>

I1 Aineistot ja metatiedot ovat sisällöltään määrämuotoisia, monikäyttöisiä, saatavilla olevia ja jaettua kieltä käyttäviä

- Datan ja metatiedon tulee olla ymmärrettävää sekä koneille että ihmisille.
- Dataa ja metatietoja täytyy voida käsitellä eri järjestelmissä ilman, että niiden merkitys muuttuu.

Ohjeita soveltamiseen tutkijalle

- Huomioi datanhallintasuunnitelmassa jo tutkimuksen suunnitteluvaiheessa tiedostomuodon käytettävyys ja datan rakenne huomioiden aineistotyyppiä, menetelmäsi ja oman tieteenalasi ja tarvittaessa lähialojen käytännöt.
 - Käytä datan keräämisessä, luomisessa ja julkaisemisessa ensisijaisesti oman alasi standardeja. Ellei näitä ole tai niitä ei voi hyödyntää on erityisen tärkeää tehdä seuraavat asiat:
 - *Dokumentoi* kunkin mittauksen, muuttujan, haastattelun tms. tarkoitus, merkitys, merkintätapa, koodaus yms. Muuten lämpötilaa ei erota euroista, kilogrammoja kilometreistä, haastateltavan ammattiluokan koodausta asuinpaikkakunnasta jne.
 - *Identifioi* käytetyt lähteet yksiselitteisesti, hyödyntäen auktoriteettipalveluja (esim. ORCID tai viralliset nimistöt).
 - *Tallenna* informaatio mahdollisuuksien mukaan vain kerran ja systemaattisesti yhdenmukaisesti samassa muodossa. Hyödynnä hierarkioita. Tee jatkoanalyysit tämän tiedon perusteella.
- Suosi avoimia, dokumentoituja ja yleisesti käytettyjä metadataformaatteja ja tiedostomuotoja aina kun mahdollista.
 - Oman tieteenalan formaatit voivat erota aineiston jakamiseen sopivista formateista.
 - esimerkiksi pitkäaikaissäilytyksen hyväksymät siirtoformaatit
 - Myös käytetty tiedostomuoto tulee ilmaista metadatatassa.

Ohjeita soveltamiseen datanhallinnan asiantuntijalle

- Tue virallisten ja *de facto* standardien löytämistä, käyttämistä ja noudattamista.
- Tue kattavaa kontrolloitujen sanastojen ja ontologioiden käyttämistä, täydentämistä ja kehittämistä.
- Hyödynnä ja tarjoa edelleen hyviä työkaluja semanttisen yhteentoimivuuden tueksi (esimerkiksi sanasto- ja ontologiapalvelut, alakohtaiset datapalvelut).



Ohjeita soveltamiseen palveluiden tuottajalle

- Huomioi käyttäjiesi tarpeet ja tieteenalojen käytännöt metadatatassa mahdollisimman laajasti ja datassa mahdollisimman tarkasti. Sanastoja voi käyttää rinnan useita:
 - suomalaisia/suomenkielisiä
 - kansainvälisiä
 - yleisiä
 - tieteenalakohtaisia
- Palvelun/palveluntarjoajan/taustajärjestelmän pitää pystyä tukemaan tutkijaa tuottamaan kuvailutietoja, joiden perusteella voidaan edelleen tuottaa semanttisesti yhteensopivaa metadataa.
- Vastaa käyttäjien tarpeeseen ottaa eri alojen sanastoja käyttöön esimerkiksi *ad hoc* rajapintojen avulla tai kopioimalla referenssimetadatoja palveluun. Sen lisäksi käyttäjälle voi myös tarjota mahdollisuutta lisätä muita termejä ja niiden tunnisteita termille tai sanastolle.
- Koneluettavuus ja semanttinen yhteentoimivuus ovat oleellisia ja käytössä sanastojen on oltava myös ihmisen ymmärrettävissä. Riippuvuuksien ja relaatioiden pitää olla ihmiselle näkyviä.
- Julkaise palvelun sisäiset sanastot FAIR-periaatteita noudattaen. Versioi ne ja linkitä niitä muihin mahdollisuuksien mukaan.
- Palveluntarjoajan tulisi käyttää tarjoamissaan sanastoissa jotain linkitetyn datan skeemaa (esim. RDF tai JSON LD).
- Sanastojen tulee olla koneluettavia. Tieteenalakohtaiset sanastot tulee mahdollisuuksien mukaan ontologisoida ja ylläpitää kytkettynä ontologiapalveluun (esimerkiksi OBO Foundry, AgrO tai FINTO).

12. Aineistoissa ja metatiedoissa käytetään sanastoja, jotka noudattavat FAIR-periaatteita

- Suosi olemassa olevia, avoimia ja koneluettavia sanastoja, ontologioita ja koodistoja.

Ohjeita soveltamiseen tutkijalle

- Mitä rikkaampi kuvailu, sen parempi, muista siis kuvailun rikastaminen pitkin matkaa.
 - Kuvaile aineisto itsellesi, siten se on ymmärrettävämpi myös muille.
 - Älä tyydy minimikuvailuun, vaan huomioi oman alan terminologia, yleisesti käytetyt termit ja lyhenteet. Valitse mahdollisimman tarkat kuvailutermit.
- Mitä paremmin aineisto on projektin aikana kuvailtu, esimerkiksi koodikirjoja käyttämällä, sen helpompaa sille on laatia löytyvyyttä ja yhteensopivuutta tukeva metadata.
- Olemassa olevan julkaistun sanaston käyttäminen, esimerkiksi paikannimistön tai lajinimistön, tekee aineistosta välittömästi yhteentoimivamman. Muita samalla sanastolla kuvattuja aineistoja voi koneellisesti verrata yhdenmukaisesti. Tämä pätee myös metadatan tasolla.



- Huomioi jo suunnitteluvaiheessa, että palvelu, jossa julkaiset aineistosi kuvauksen, voi vaikuttaa siihen mitä sanastoja on tarjolla.
- Oman alan sanastojen tunteminen on osa tutkijan ammattiosaamista. Tämä tarkoittaa esimerkiksi soveltuvien sanastojen valintaa siten, että tiedot voi yhdistää muihin saman alan vastaaviin tietoihin kansainvälisesti.
- Pyydä tarvittaessa tukea sopivan sanaston löytämiseen ja koulutusta sen käyttöön oman organisaation datatuesta, kirjastosta, oman alan datapalveluilta tai muilta tutkijoilta. Jaa tietoa.
- Anna palautetta, jos sanasto ei ole riittävä tai kaipaa muutoksia.

Ohjeita soveltamiseen datanhallinnan asiantuntijalle

- Tunne keskeisimmät yleiset sanastot sekä semanttiset hakupalvelut ja ohjaa tarvittaessa kunkin erikoisalan sanastopalveluiden äärelle.
- Tunne sanastojen ja pysyvien tunnisteiden erot, merkitys ja käyttö seuraavissa yhteyksissä:
 - tutkimusaineistojen kuvailu ja sisällön ymmärrettävyys
 - Tutkimusaineistojen linkittyminen toisiinsa kuvailutermin tai niiden tunnisteiden välityksellä. Kuvailutermit auttavat hakemaan ja löytämään samankaltaisia aineistoja ja/tai muita tutkimustuotoksia.
- Huomioi ja hyödynnä lähituessa tieteenalakohtaisten sanastojen tuntijoita.
 - Ota selvää sopivista tieteenalakohtaisista sanastoista.
 - Keskustele tutkijoiden ja muiden asiantuntijoiden kanssa.

Ohjeita soveltamiseen palveluiden tuottajalle

- Käytä avoimia, versioituja, ylläpidettyjä, alakohtaisia ja/tai yleisiä sanastoja ja ontologioita.
 - Avoimien sanastojen ja ontologioiden tulisi olla itsessään FAIR, eli koneluettavia ja dokumentoituja.
 - Uusia sanastoja ja tietomalleja rakennettaessa tulee kiinnittää huomiota, että ne julkaistaan ja kuvataan FAIR-periaatteita noudattaen.
 - Sanastot tulee toteuttaa niin, että niiden käyttäjät voivat osallistua niiden ylläpitoon esimerkiksi ehdotus- ja korjaustoimintojen avulla.
- Julkaise omat referenssidataksi kelpaavat sanastot itse tai jossain palvelussa.
 - esimerkiksi Luomus on julkaissut suomenkieliset eliönimistöt eri aikakausina ja Maanmittauslaitos ylläpitää tietokantoja kuntien nimistä ja rajoista
- Sanastojen ja niiden käsitteillä tulisi olla pysyvät tunnisteet ja niiden tulisi olla käytettävissä avoimien rajapintojen kautta.
- Hyödynnä käyttöä helpottavia teknologioita, kuten esimerkiksi tekoälyyn pohjautuvaa suosittelua aineistojen kuvailun yhteydessä ja metadatan rikastamiseen.
- Huolehdi metatietojen kuratoinnista myös pitkällä aikavälillä.



13. Aineistoissa ja metatiedoissa on tyypiteltyjä viittauksia muihin resursseihin

- Tyypittely tarkoittaa, että viittauksen luonne on määritelty; esim. A on B:n aiempi versio tai B on käyttänyt A:ta lähteenä.
- Pysyvien tunnisteiden käytöllä voidaan turvata viittausten kestävyys pitkällä aikavälillä.

Ohjeita soveltamiseen tutkijalle

- Muista, että semanttisen yhteentoimivuuden huomioiminen on osa vastuullista tiedettä, joka koskee kaikkia tieteenaloja, vaikka datasi ei olisi avointa.
 - Ole tietoinen ja ota selvää semanttisesta yhteentoimivuudesta.
 - Selvitä mistä saa apua tallennuspaikan eli repositorion valintaan (datatuki, data-agentit, tutkijakollegat, repositorio).
 - Valitse jos mahdollista sellainen tallennuspaikka datalle, joka tarjoaa linkitetyn datan palveluita, jolloin aineistot ja käytetyt sanastot löytyvät hakupalveluiden avulla.
- Vaadi käyttäjäystävällisiä ja ymmärrettäviä palveluita, jotka
 - laativat aineistolle yksityiskohtaiset kuvailut.
 - tarjoavat pysyvät tunnisteet ja sanastot.
 - mahdollistavat rajapinnat (meta)datan hyödyntämiseen.
- Dokumentoi tutkimuksesi riippuvuudet huolella.
 - Metadatatassa tulisi olla viittauksia sekä eri data-aineistojen välille että muuhun metadataan.
 - Metadataan tulisi laatia tunnisteita ja yksilöiviä käsitteitä käyttäen viittaukset eri data-aineistojen välille sekä muiden aineistojen metadataan.
- Muista versionhallinta.

Ohjeita soveltamiseen datanhallinnan asiantuntijalle

- Kouluttajan on tunnettava perusteet linkitetyn datan vaatimuksista ja vaikutuksista sekä ymmärrettävä tieteenalakohtaiset näkökulmat.
 - Dataan viittaaminen ja elinkaaren kuvaaminen tukevat toistettavuutta.
 - Linkitetty data lisää tutkijan ja tutkimuksen näkyvyyttä ja vaikuttavuutta sekä datan uudelleenkäyttöä.
 - Aineistoja hyödyntävien laskennallisten algoritmien ja työvirtojen avoimuus ja saavutettavuus kasvattavat osaltaan aineistojen arvoa ja edistävät niiden jatkokäyttöä. Aineiston yhteyteen voi lisätä viitteitä aineiston hyödyntämistä tukeviin menetelmiin ja käyttöesimerkkeihin.
- Sanastot ja tietomallit kuvastavat tieteenalan ontologiaa, varmista aina näiden sopivuus tutkijoilta.
- Huomioi metadatan tarkistus ja täydennys koko tutkimusprosessin aikana (kuratointi).



- Jos tiedetään, että tutkimusaineisto on pitkäaikaissäilytettävää, datanhallinnan asiantuntijan tulee pystyä neuvomaan, miten pitkäaikaissäilytyksen metatietojen vaatimukset huomioitaisiin jo alusta asti. Näihin kuuluvat kaikkia aineistotyyppisiä koskien tiedostoformaattit ja niiden versiot, formaattien dokumentointi, tarkistussummat ja muut tutkimusaineiston tulkinnan erityisvaatimukset.

Ohjeita soveltamiseen palveluiden tuottajalle

- Tunnista kohdeyhteisön tarpeet ja käytännöt ja seuraa niitä. Huolehdi vuorovaikutuksesta ja sisällön ja ratkaisujen kehittämisestä.
- Käytä kansallisten ja kansainvälisten keskeisten tietovarantojen tuntemia pysyviä tunnisteita, jos mahdollista.
- Huomioi pysyvien tunnisteiden ja sanastojen kohdalla niiden lisenssi, laatu ja luotettavuus.
- Käytä kontrolloituja arvoja (esim. sanaston käsite tai toimijan tunniste) ja mahdollisuuksien mukaan pysyviä tunnisteita mahdollisimman paljon.
 - Aineistoa kuvailevassa metadatasissa: Lisätään käsitteisiin pysyviin tunnisteisiin perustuvia linkkejä muihin tietojärjestelmiin (esim. finto.fi).
 - Hyödynnä myös organisaatiolle, tekijöille ja muille toimijoille tarjolla olevia tunnisteita (esim. ORCID). Edistä kansallisten resurssien kehittämistä.
 - Rakennetta kuvailevassa metadatasissa: Metadatan tulisi koneluettavasti kuvata datan rakennetta. Käytä esimerkiksi DataCiten metadataskeemassa listattuja relaatioita kuvaamaan esimerkiksi tietoaineistojen välisiä suhteita, esimerkiksi versioita.
 - Hallinnollisia ominaisuuksia kuvailevassa metadatasissa: Lisenssit, käyttöoikeuksien rajoitukset ja niiden perusteet.
- Tavoittele käyttäjätavallisia palveluita, joissa on integroituina toiminnallisuuksina luokitteluja, sanastoja ja pysyviä tunnisteita. Linkitysten ja koneluettavuuden pitäisi käyttäjän kannalta syntyä automaattisesti.
- Referenssimetadatasia käyttämällä voi rajata sallittuja arvoja ja edistää koneluettavuutta. Laajennuksia ja vapaatekstiä on myös hyvä sallia, mutta omilla kenttissään.



Laatijat

Avoimen tieteen koordinaatio, Fair-periaatteiden soveltaminen -työryhmä

Alaterä, Tuomas; <https://orcid.org/0000-0002-3448-3448>
Asmi, Ari; <https://orcid.org/0000-0003-3933-4684>
Borg, Maisa; <https://orcid.org/0000-0003-2215-8303>
Fält, Katja; <https://orcid.org/0000-0002-6172-5377>
Hakala, Juha; <https://orcid.org/0000-0003-1067-5020>
Hakkarainen, Jussi-Pekka; <https://orcid.org/0000-0002-9055-9312>
Holappa, Anne; <https://orcid.org/0000-0003-3044-0173>
Huuskonen, Saira; <https://orcid.org/0000-0002-2316-285X>
Koikkalainen, Riitta; <https://orcid.org/0000-0003-3289-1832>
Koivula, Hanna; <https://orcid.org/0000-0001-5605-9122>
Kuuskoski, Reetta
Kuusniemi, Mari Elisa; <https://orcid.org/0000-0002-7675-287X>
Kylmälä, Kaisa; <https://orcid.org/0000-0001-8227-6817>
Laihonen, Petteri; <https://orcid.org/0000-0002-3914-0954>
Laine, Katja; <https://orcid.org/0000-0003-3820-5899>
Lehto, Anttoni; <https://orcid.org/0000-0002-6502-5351>
Lindholm, Tanja; <https://orcid.org/0000-0001-6115-3359>
Marttila, Juuso; <https://orcid.org/0000-0003-3394-0341>
Mäkinen, Tarja; <https://orcid.org/0000-0001-7658-6026>
Nurmi, Niina
Nygren, Pekka; <https://orcid.org/0000-0002-3730-9840>
Näppilä, Turcka; <https://orcid.org/0000-0002-0562-7254>
Pajari, Jussi; <https://orcid.org/0000-0002-3843-1442>
Parland-von Essen, Jessica; <https://orcid.org/0000-0003-4460-3906>
Pääkkönen, Tuula; <https://orcid.org/0000-0003-3958-9732>
Päällysaho, Seliina; <https://orcid.org/0000-0002-3554-7773>
Salminen, Nina-Mari; <https://orcid.org/0000-0002-1864-9428>
Sipola, Tiina; <https://orcid.org/0000-0002-9242-8463>
Sipponen, Sonja; <https://orcid.org/0000-0002-4953-6354>
Söderholm, Maria; <https://orcid.org/0000-0002-7100-547X>
Taskinen, Timo; <https://orcid.org/0000-0003-3538-0720>
Troberg, Heidi
Vuorinen, Marjut; <https://orcid.org/0000-0003-1919>