



## Missä julkaisen TKI-datan?

7.5.2024



# Webinaarin sisältö

Keskeisten termien määritelmät

Mitä ottaa huomioon, kun valitset datalle julkaisupaikkaa?

Esimerkkejä datarepositorioista ja niiden ominaisuuksista

*Webinaari tallennetaan. Kysymyksiä voi esittää chatin kautta tai kysyä ja kommentoida webinaarin kysymyssessioiden aikana nostamalla käden ja pyytämällä puheenvuoroa. Tallenne ja diat tulevat:*

*<https://tt.eduuni.fi/sites/amkit/avoin/SitePages/Webinaarit%202024.aspx>*

## Termi: Tutkimusaineisto

- Webinaarissa keskitymme **digitaaliseen tutkimusaineistoon**, eli **dataan**, joka sisältää tyypillisesti tiedostoja.
- Datan liitteeksi tarvitaan kuvailevaa ja teknistä tietoa siitä mitä informaatiota se sisältää. Yhdessä datan kanssa tästä kokonaisuudesta muodostuu **tutkimusaineisto**.

### Tutkimusaineisto

*Tutkimusaineisto eli tutkimusdata on tutkimuksen aikana itse kerättyä tai hyödynnettyä aineistoa, johon tutkimustulokset pohjautuvat.*



# Termi: Pysyvä tunniste (persistent identifier, PID)



DOIs include a prefix (prefixes always start with 10.) and a suffix, separated by a forward slash (/). Prefacing the DOI with [doi.org/](https://doi.org/) will turn it into an actionable link, for example, <https://doi.org/10.47366/sabia.v5n1a3>. Clicking that link will 'resolve' it, i.e. redirect to the latest information about the object it identifies, even if the object changes or moves.

- **Pysyviä tunnisteita** ovat esimerkiksi **DOI** ja **URN**, joita käytetään tutkimusaineistoille, ja tutkijatunniste **ORCID**. Pysyvien tunnisteiden käyttö on yksi FAIR-periaatteiden kulmakivistä.
- **Pysyvä tunniste on viite**, jonka avulla **tutkimusaineisto voidaan aina löytää**, vaikka aineiston sijainti tai palvelut muuttuisivat. Pysyvää tunnistetta hallinnoiva repositorio vastaa siitä, että pysyvä tunniste vie aina oikealle kuvailusivulle. Pysyvä tunniste vie myös aina tutkimusaineiston kuvailusivulle, vaikka itse data ei enää olisi saatavilla (ns. muistosivu)
- Kun TKI-datallasi on **pysyvä tunniste**, sinun ja muiden tulee käyttää sitä dataan **viitattaessa**. Näin kaikki viittaukset voidaan yhdistää aineiston tekijöihin.
- [https://tieteentermipankki.fi/wiki/Avoin\\_tiede:pysyv%C3%A4\\_tunniste](https://tieteentermipankki.fi/wiki/Avoin_tiede:pysyv%C3%A4_tunniste)

## Termi: FAIR-periaatteet

- FAIR-lyhenne tulee sanoista:
  - Findable, Accessible, Interoperable ja Re-usable.
  - Näitä kutsutaan FAIR-periaatteiksi
- Periaatteet tarkoittavat, että tutkimuksessa käytettyjen lähdeaineistojen, menetelmien ja tietotuotteiden tulee olla **löydettävissä, saatavilla, siirrettävissä tai yhdistettävissä ja uudelleenkäytettävissä.**  
(<https://www.opendata.fi/fi/opas/tietoaineiston-fair-periaatteet>)
- FAIR ja Open Science:
  - "niin avointa kuin mahdollista, niin suljettua kuin välttämätöntä"




Esimerkiksi monet tutkimusrahoittajat ohjaavat noudattamaan FAIR-periaatteita



# FAIR-periaatteet käytännössä

- Data kuvaillaan hyvin.
- Koneluettavaa ja hyvin jäsenettyä (meta)tietoa, esim. asiasanat
- Metatiedot kaikkien löydettävissä
- Lisensoitua
- Pysyviä tunnisteita sisältävää
- Aineistot voi linkittää esim. rahoitukseen, organisaatioihin ja julkaisuihin
- FAIR on hyvä tavoite!

<b>F</b>	<b>FINDABLE – LÖYDETTÄVÄ</b> <ul style="list-style-type: none"><li>• Olennaiset tiedot kuvailtu riittävän tarkasti</li><li>• Aineistolla kuvailusivu ja ainutkertainen pysyvä tunniste</li></ul>	
<b>A</b>	<b>ACCESSIBLE – SAAVUTETTAVA</b> <ul style="list-style-type: none"><li>• Voidaan hakea internetistä</li><li>• Versiointi ja elinkaaren dokumentointi</li><li>• Äineistosta löytyy muistosivu, vaikka data olisi poistettu</li></ul>	
<b>I</b>	<b>INTEROPERABLE – YHTEENTOIMIVA</b> <ul style="list-style-type: none"><li>• Käytössä yleiset, dokumentoidut ja avoimet tiedostomuodot</li><li>• Yhteisten koodistojen käyttö &amp; linkitykset muihin tuotoksiin</li></ul>	
<b>R</b>	<b>RE-USABLE – UUELLEENKÄYTETTÄVÄ</b> <ul style="list-style-type: none"><li>• Data on hyvin dokumentoitu ja ymmärrettävää</li><li>• Käyttöoikeudet/lisenssit selkeästi esillä</li></ul>	

## Termi: Datarepositorio



*Tutkimusaineistojen tallentamiseen, jakamiseen, etsimiseen ja hyödyntämiseen tarkoitettu palvelu.*

- Datarepositorioon tallennetaan varsinaisen tutkimusaineiston tiedostojen lisäksi sen **kuvailutietoja**.
- Suositeltavat datarepositoriot myöntävät tutkimusaineistoille **pysyvät tunnisteet** ja vastaavat niiden toimivuudesta.



# Kysymyksiä





## Miksi datarepositorioita tarvitaan?



FAIR-periaatteita ei pysty toteuttamaan ilman palveluja.

Datarepot tarjoavat pysyvän tunnisteiden ja näin linkit eivät mene rikki, vaikka verkko-osoitteet muuttuvat. Lisäksi dataan voidaan viitata.

Kun dataan viitataan, nimesi löytyy aina datan yhteydestä; meritoidut.

Aineisto säilytetään palvelussa, eikä esim. omalla koneella.

Monet repositoriot mahdollistavat aineiston linkittämisen organisaatioon.

## Datarepositorioita on erilaisia

- Ei ole olemassa yhtä kaiken kattavaa kriteeristöä datarepositoriolle
- Datarepositorioilla voi olla erilaisia **kypsyystasoja** ja **ominaisuuksia**, kuten:
  - Miten repositorio ottaa huomioon koneluettavuuden tai mahdollistaa FAIR-periaatteiden toteuttamisen
  - Miten aineistosta huolehditaan sen julkaisemisen jälkeen
  - Miten tekninen säilyttäminen tai jakelu on toteutettu
  - Metatietojen siirtyminen / haravointien mahdollisuus
- ***Tarkista kotiorganisaatioltasi, miten tieto aineistosi olemassaolosta voidaan kerätä***

Kansallisesti tutkimusaineistojen julkaisutietoja kootaan **Tiedejatutkimus.fi-portaaliin**, mutta kaikista datarepositorioista ei siirry sinne automaattisesti tietoja.

**Tiedejatutkimus.fi-palveluun siirtyy tällä hetkellä tietoja seuraavista repositorioista:**

Fairdata, Tietoarkisto, Kielipankki sekä tiettyjen organisaatioiden omat repositoriot ja metatietovarannot.

# Datarepositorion valinta – Suunnittele ajoissa ja tutki vaihtoehtoja

- On tärkeää suunnitella jo tutkimuksen alussa:
  - o Mikä osa aineistosta halutaan julkaista?
  - o Missä datarepositoriossa aineiston voi julkaista?
    - Tämä auttaa varautumaan ennakolta mm. tiedostoformaatti- ja metadatatavaatimukseen, tietosuojakysymyksiin ja suunnittelemaan aineiston julkaisemisen vaatimia resursseja.
- Metadatan voi julkaista monessa eri repositoriossa, myös pelkän metadatan voi julkaista.
- Ennalta varautumalla minimoit riskit ja saat tuloksista täyden hyödyn irti.

## Datarepositorion valinta – Yleisohjeita

- Mieti datan hyödyntäjää:
  - Mistä data löytyy hyödyntäjälle helpoimmin?
  - Mitkä ovat palvelun haku- tai filteröintimahdollisuudet eli miten helposti dataa voi palvelusta hakea?
  - Onko tieteenalasi näkökulmasta tieteenalakohtainen repositorio datan löydettävyyden näkökulmasta kätevä?
- Kysy apua oman organisaation tukipalveluista ja tutustu kotiorganisaatiosi ja rahoittajan ohjeisiin. Sopivia datarepositorioita voi hakea oman organisaation sivuilta ja esim. <https://www.re3data.org/> -palvelusta.

### Tieteenalakohtaisten palvelujen etuina voivat olla:

- tieteenalakohtaiset metatietokentät ja muuttujat
- datan visualisointimahdollisuudet (esim. karttapalvelu)
- yhteentoimivat tiedostoformaattit
- aineiston löydettävyys.

*Lisäksi yhteistoiminta muiden saman alan toimijoiden kanssa.*

# Näkökulmia datarepositorion valintaan – Jatkuvuus ja laatu

## Säilyvyys



Säilytysaika; mitä luvataan?  
Säilytetäänkö data palvelussa  
useampana kopiona?  
Voidaanko aineisto poistaa  
tekijöiden tai kotiorganisaation  
tietämättä?

## Repositorion jatkuvuus



Kuka palvelusta vastaa?  
Onko kyseessä luotettava,  
vakavarainen toimija?

## Repositorion käyttöehdot

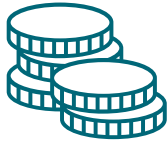


Tarkista palvelun käyttöehdot ja  
politiikat:

Ovatko vastuut ja roolit selvillä?  
Siirtyykö oikeuksia tai  
rekisterinpitäjyys eri taholle datan  
datarepositorioon siirron  
yhteydessä?

# Näkökulmia datarepositorion valintaan - Käytettävyys

## Maksullisuus



Ottaako repositorio maksun datan tallennuksesta? Riippuuko se esim. datan määrästä?

## Tuki



Tarjoaako repositorio tukea tai ohjeita aineiston luomiseen?

esim. anonymisoinnin tarkistus  
→ Tietoarkisto

## Helppokäyttöisyys



Osaatko käyttää palveluita?  
Onko tarjolla oppaita?

# Näkökulmia datarepositorion valintaan - Käytettävyys

## Aineistosta huolehtiminen



Mikä taho hallinnoi aineistoa sen julkaisun jälkeen (kuratointi)?  
Ylläpidetäänkö aineiston **ymmärrettävyyttä ja käytettävyyttä**?  
Tiedostomuodot säilyvät käytettävänä.  
Jos aineistoa jaetaan rajoitetusti, kuka tätä hoitaa?

## Aineiston käyttöehdot



Onko aineistolle mahdollista asettaa uudelleenkäytön mahdollistava lisenssi?

[Creativecommons.fi](https://creativecommons.org/)

# Näkökulmia datarepositorion valintaan – Ominaisuudet

## Aineiston metatiedot



Tarkista repositorion kuvailukenttien minimimäärä/riikkaus.  
Tarvitsetko tieteenalakohtaista kuvailumahdollisuuksia?  
Voitko linkittää aineistoon esim. menetelmiä tai julkaisuja?

## Haku



Onko aineisto helppo löytää hakupalvelujen avulla?

## Kasuvat aineistot / dynaaminen data



Onko sinulla tarvetta lisätä aineistoon dataa jälkikäteen?  
Tuetaanko tätä?



# Näkökulmia datarepositorion valintaan – Ominaisuudet

## Jakelu



Onko mahdollisuus **avata** aineisto avoimesti, **luvittaa** pääsy dataan tai jakaa dataa rajoitetusti?  
Voiko datalle määrittää **embargon**?

## Datan määrä, koko & rakenne



Millaista dataa palvelu ottaa vastaan ja missä muodossa sitä jaellaan?

## Raportointi



Onko mahdollista seurata aineiston katselu- ja latauskertoja?

## Näkökulmia datarepositorion valintaan – Henkilötiedot

Jos datasi sisältää henkilötietoja, suunnittele tutkimuksen jälkeinen säilytys jo ennen kuin keräät dataa. Kysy apua kotiorganisaatioltasi ja sen tietosuojavastaavalta

Hyviä ohjeita löydät Tietoarkiston sivulta. Mm.

<https://www.fsd.tuni.fi/fi/palvelut/aineistonhallinta/tutkittavien-informointi/>

# Eriyisten henkilötietoryhmien käsittely

- Tällaisista tiedoista ilmenee henkilön
  - rotu tai etninen alkuperä
  - poliittisia mielipiteitä
  - uskonnollinen tai filosofinen vakaumus
  - ammattiliiton jäsenyys
  - terveyttä koskevia tietoja
  - seksuaalinen suuntautuminen tai käyttäytyminen
  - geneettisiä ja biometrisia tietoja henkilön tunnistamista varten.
- Tietoja voi käsitellä mm. seuraavissa tapauksissa:
  - Kun rekisteröity on antanut **nimenomaisen suostumuksen** kyseisten henkilötietojen käsittelyyn.
  - Kun käsittely on tarpeen **yleisen edun mukaista arkistointia, tieteellistä ja historiallista tutkimusta tai tilastointia varten** tietosuojasetuksen mukaisesti unionin oikeuden tai jäsenvaltion lainsäädännön nojalla.

Näitä tietoja on suojeltava erityisen tarkasti, koska niiden käsittely voi aiheuttaa huomattavia riskejä henkilön perusoikeuksille ja -vapauksille.

<https://tietosuoja.fi/eriyisten-henkilotietoryhmien-kasittely>

# Näkökulmia datarepositorion valintaan – Henkilötiedot

- 1. Varmista käsitteletkö erityisiä henkilötietoja
- 2. Mikäli käsittelet, onko sinulla erityinen syy?
- 3. Mikäli on, selvitä etukäteen ohjeet kotiorganisaatiostasi/tietosuojavastaavalta/Tietoarkistosta (Aineistonhallinnan käsikirja)
- 4. Vaikka saisit luvan käsitellä henkilötietoja, niitä ei saa koskaan avoimesti julkaista sellaisenaan (esim. ilman anonymisointia).
- Usein datan vienti datarepositorioon ei muuta rekisterinpitäjää. Tällöin tyypillisesti tutkija tai kotiorganisaatio on **rekisterinpitäjä**, repositorio puolestaan henkilötietojen käsittelijä. Tarkista tämä kuitenkin erikseen.
  - Aineiston rekisterinpitäjällä on viime kädessä vastuu aineiston käsittelystä lakien ja muiden velvoitteiden mukaisesti.

# Kuinka valita erityisesti kansainvälisille hankkeille datarepositorio, johon tallentaa data?

- Oleellisia asioita tarkastella:
  - Rahoittajan ja kotiorganisaation linjaukset
  - Yhteistyömahdollisuudet, esim. saako kaikille osapuolille tunnukset palveluun, onko mahdollista tallentaa ja kuvailla dataa yhteistyössä
  - Metadatat löydettävissä
  - Tieteenalakohtaiset mahdollisuudet
  - Kansallinen palvelu voi olla usein oikein hyvä vaihtoehto myös kansainvälisten projektien datarepositorioksi

# Esimerkkejä datarepositorioiden ominaisuuksista



Repositorion nimi	Tieteenala	Yksittäisen aineiston koko	Maksullisuus	Pysyvät tunnisteet
Zenodo	Monitieteinen	max 50 GB	maksuton	Kyllä (DOI)
Dryad	Monitieteinen, etenkin bioaineistoille	Yksittäinen tiedosto max 10 GB, aineisto 300 GB	Maksullinen, 150 \$ per julkaisu	Kyllä (DOI)
Fairdata (IDA, Qvain, Etsin)	Monitieteinen	max ~ kymmeniä TB	maksuton	Kyllä (DOI, URN)
Paituli	Paikkatietodata	max ~2-3 TB	maksuton	Kyllä (URN Fairdata Etsimessä)
Kielipankki	Kielitieteet	max ~250 GB	maksuton	Kyllä (URN, Handle)
Tietoarkiston Aila	Kaikki, erityisesti yhteiskunta-, humanistiset ja kasvatustieteet	Ei tarkkaa kokorajaa, sovittava Tietoarkiston kanssa	maksuton	Kyllä (URN, tulossa DOI)
EUDAT CDI B2SHARE / EUDAT B2SHARE Premium	Monitieteinen ja/tai yhteisökohtainen	CDI B2SHARE: max 20 GiB / aineisto Premium-palvelut: räätälöitynä	CDI-palvelut: Maksuttomia Premium-palvelut: maksullisia (asiakasorganisaatiolle)	Kyllä (DOI, Handle)



# Kysymyksiä

# Mikä on EOSC? Voiko EOSCissa julkaista dataa?





## Visiosta toteutukseen “EOSC federaatio”

- EOSC:n visiona on luoda datan ja palvelujen verkko eurooppalaista tiedettä varten.
- Keskeinen osa tätä tavoitetta on Euroopassa olemassa olevien infrastruktuurien luotettava virtuaalinen yhteentoimivuus FAIR-periaatteiden mukaisten tutkimustulosten tallentamiseksi, jakamiseksi ja uudelleenkäytöksi yli rajojen ja tieteenalojen, jota kutsutaan myös EOSC-federaatioksi.
- Federaatio tarjoaa koordinoitua ns. solmukohtia/yhteyspisteitä ensisijaisesti eurooppalaisille tutkijoille, jotka voivat käyttää FAIR-dataa ja yhteentoimivia palveluja tutkimusprosessin aikana. Yhteyspisteet tulevat olemaan samoja, joita tutkijat käyttävät nykyisinkin; esim. kansalliset tai temaattiset infrastruktuurit.

**1. ...a cloud infrastructure**

Despite the word "cloud" part of its name, EOSC is not a new cloud computing platform

**2. ... a new research data repository or research data management system**

The federation of existing infrastructures, i.e. EOSC, is a new infrastructure which does not exist today

**3. ... a new pan-European e-infrastructure**

EOSC is not building a new e-infrastructure. EOSC is building i. the components to enable the federation of existing data, research and e- infrastructures nodes and ii. the additional services needed to enable the Web of FAIR data and related services.

**4. ... synonymous of Open Science**

EOSC is the enabler that will support the deployment of Open Science in Europe. EOSC does not substitute any existing Open Science networks.

**5. ... the EOSC Association**

The EOSC Association as representative of the various stakeholders in Europe is the legal entity established to work together with the European Commission to support the realisation of the EOSC strategy.

**6. ... substituting any existing national, regional, pan-European, agnostic nor thematic Research Infrastructures or e- infrastructures**

EOSC will enable the federation of existing data, research and e-infrastructure nodes. The new developments are focused on components enabling the federation and on the additional services needed to enable the Web of FAIR data and related services

**7. ... the EOSC Portal**

The EOSC Portal is one of the results of the EOSC Future EC funded project (2019-2023). The EOSC Portal is piloting the EOSC AAI and the idea of a European marketplace for services supporting researchers.

**8. ... owning any data or services**

EOSC is an enabler. The ownership of the federated elements (data, services, research infrastructures, e-infrastructure, etc.) will remain with the providers.

**9. ... engaging directly individual researchers.**

Individual researchers will benefit from EOSC through their existing channels (e.g. universities, research institutes, research infrastructures, associations, etc.) that will act as intermediaries.

# Rahoittajien vaatimuksia datan julkaisemisen suhteen

- Avoin tiede ja tutkimus ovat **Suomen Akatemian** strategisia tavoitteita, ja yksi rahoituksen perusteista. Suomen Akatemian tavoitteena on, että tutkimuksen tulokset ja metatiedot ovat nopeasti ja helposti saatavilla jatkokäyttöä varten.
- **Business Finland** on sitoutunut avoimen tieteen ja tutkimuksen tavoitteisiin. Näihin kuuluu, että julkisin varoin tuetuissa organisaatioissa tuotetut julkaisut ja tieto, joka ei ole salassapidettävää, olisi avoimesti saatavilla jatkokäyttöä varten.
  - <https://www.aalto.fi/fi/avoin-tiede-ja-tutkimus/rahoittajien-avoimen-tieteen-vaatimukset>

# Rahoittajien vaatimuksia datan julkaisemisen suhteen

- **Euroopan Komissio: Horizon 2020;** Nykyisten vaatimusten mukaan data tulee avata data-arkistossa, jotta tutkimustulokset on mahdollista validoida. Jos olet epävarma siitä, tuleeko sinun avata dataa vai ei, tarkista asia projektisi rahoitussopimuksesta (Grant Agreement, article 29.3 Open access to research data).
- **Horizon Europe** (vuoteen 2027 asti); sisältää pakollisia ja suositeltuja avoimen tieteen käytäntöjä. Euroopan komissio (EC) on mukana tutkimuksen avoimuutta edistävässä Plan S -aloitteessa, joka vaikuttaa avoimen julkaisemisen ja datan vaatimukseen vuoden 2021 rahoitushauista alkaen.

○ <https://www.aalto.fi/fi/avoin-tiede-ja-tutkimus/rahoittajien-avoimen-tieteen-vaatimukset>

## Lisätietoa datan julkaisusta ja FAIR-periaatteista

- Manage your research data - towards publishing data  
[https://video.csc.fi/media/t/o\\_xoejbmmk](https://video.csc.fi/media/t/o_xoejbmmk)
- The Elements of FAIR – Findable, video [https://video.csc.fi/media/t/o\\_q2jf8roy](https://video.csc.fi/media/t/o_q2jf8roy)
- The Elements of FAIR – Accessible, video [https://video.csc.fi/media/t/o\\_z9g5gwhu](https://video.csc.fi/media/t/o_z9g5gwhu)
- The Elements of FAIR - Interoperability [https://video.csc.fi/media/t/o\\_l6kfkwm4](https://video.csc.fi/media/t/o_l6kfkwm4)
- The Elements of FAIR - Reusable [https://video.csc.fi/media/t/o\\_58184bc4](https://video.csc.fi/media/t/o_58184bc4)
- Suomenkielinen tutkimusdatanhallinnan kurssi: <https://e-learn.csc.fi/>  
hakusana "Tutkimusdatanhallinta"
- Justuksen uusi tutkimusaineistotietojen tallennuspalvelu, keväällä tuotantoon:  
<https://wiki.eduuni.fi/display/CscJustus/Tutkimusaineistotietojen+tallennuspalvelu>



[facebook.com/CSCfi](https://facebook.com/CSCfi)



[twitter.com/CSCfi](https://twitter.com/CSCfi)



[youtube.com/CSCfi](https://youtube.com/CSCfi)



[linkedin.com/company/csc---it-center-for-science](https://linkedin.com/company/csc---it-center-for-science)



[github.com/CSCfi](https://github.com/CSCfi)