



# DATA JA LISENSSIT

25.4.2024  
Soile Manninen



# LISENSSI = KÄYTTÖLUPA

- Lisenssi voidaan rinnastaa sopimukseen
  - Onko data vapaasti saatavilla, miten dataa saa jatkokäyttää jne.
- Lisenssillä annetaan käyttöoikeus, ei hallinnointi-/omistusoikeutta
- Koskee aineistoja, jotka kuuluvat tekijänoikeuden tai tietokantasuojan piiriin
- Pitäisi antaa aina datalle, aineistoille ja ohjelmistoille, kun niitä julkaistaan tai laitetaan data-arkistoon (data repository)
- Henkilötietojen jakamiseen ei ole mitään lisenssejä!



# AVOIMET LISENSSIT

## [Avoimia lisenssejä on runsaasti](#)

- Käytetyin kansainvälinen avoin lisenssi Creative Commons (CC0, CC BY, CC BY-SA jne.),
- uusin versio 4.0 vuodelta 2013, aiemmat versiot eivät huomioi tietokantaoikeutta (EU)
- Voi käyttää lisenssinä aineistoihin, jotka suojataan tekijänoikeudella (teksti, kuvat jne.)
  - Lisenssi peruuttamaton, voimassa tekijänoikeuden mukaisesti (70 v. tekijän kuolinvuotta seuraavasta vuodesta)
- CC-lisenssillä annetaan laajemmat oikeudet jatkokäyttöön: muokkaus, kopiointi, jakaminen, kääntäminen
  - Muokkauksista pitää aina kertoa, viitata alkuperäiseen tekijään ja lisenssiin

## [Avoimelle lähdekoodille](#) omat lisenssit

- V. 2022-2023 tilastoissa eniten käytettyjä [MIT](#) ja [Apache 2.0](#), aiemmin GNU
- Esikuvina muille avoimille lisensseille



# DATA POLICY, DATA AVAILABILITY STATEMENT

- Tieteellisissä julkaisuissa on yleensä kerrottu, mistä ja miten artikkeliin liittyvä data on saatavilla
  - Artikkelissa Data Availability Statement (DAS), jonka määrittää julkaisun, kustantajan tai rahoittajan datalinjaus (Data Policy)
  - Ennakkotehtävä kuuluu: etsi tieteellinen artikkeli omalta alaltasi tai katso tieteellisen kustantajan/julkaisijan datalinjaus (esim. hakemalla: kustantajan tai julkaisijan nimi “data policy”)
    - Löytyykö artikkelista tiedot, mistä datan saa?
    - Löytyykö tietoa datan lisenssistä artikkelista vai onko mentävä datan julkaisevaan palveluun tai data-arkistoon?
    - Kustantajan / datan julkaisijan datalinjauksesta pitäisi olla suosituslisenssi – löytyykö?
    - Löytyykö muuta yllättävää tai uutta?



# ESIMERKKEJÄ: KUSTANTAJAT JA DATAJULKAISIJAT

Artikkeliin liittyvä data - kaupallinen kustantaja (UK):

<https://www.cambridge.org/core/services/authors/open-data/data-availability-statements>

Artikkeliin liittyvä data - OA-julkaisija (California, US):

<https://journals.plos.org/plosone/s/data-availability>

Datajulkaisija - DRYAD (North Carolina, USA): <https://datadryad.org/stash/mission>

“Dryad publishes data exclusively under a Creative Commons Public Domain License (CC0) and does not support the publication of sensitive data to which access should be restricted.”



# DATAN AVOIN SAATAVUUS – MUISTILISTA JULKAISUIHIN

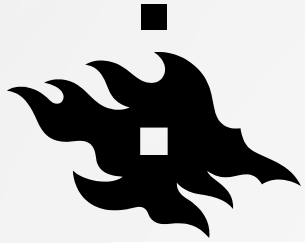
- Lehden/julkaisijan/rahoittajan datalinjauksen (data policy) mukainen
  - Lähtökohtana avoimuus / saatavuus, mutta lainsäädännöllisistä tai tutkimuseettisistä syistä linjauksesta voi poiketa
  - Jos poiketaan datalinjauksesta, kerro perustelu
    - Esim. data sisältää henkilötietoja tai tekijänoikeus rajoittaa aineistojen jakamista
  - Jos tutkimusdataa on avattu tai sen voi jakaa, kerro miten datan löytää (mm. verkko-osoite tai yhteyshenkilö).
- Käytä datalle / aineistolle pysyvää tunnistetta (esim. DOI)
- Datalla on yleensä eri lisenssi kuin artikkelilla – hyvä tehdä selväksi itselle, artikkelin lukijalle ja datan käyttäjälle!



# LAINSÄÄDÄNTÖ

## Avoimen datan direktiivi (1024/2019/EU)

- sovelletaan mm. julkisin varoin tuotettuihin tutkimusaineistoihin, jotka on julkaistu sähköisessä muodossa.
- tutkimusaineistojen oltava saatavilla ”eri esitysmuodoissaan, kielitoisintoina ja maksutta”
- suositellaan avoimia, standardoituja lisenssejä datan jakamiseen
- arvokkaat aineistot määritelty erikseen
- ei koske ohjelmistoja
- Suomessa [Laki julkisin varoin tuotettujen tutkimusaineistojen uudelleenkäytöstä 713/2021](#)
- [Avoindata.fi](#), tiedot siirtyvät [EU:n avoimen datan portaaliin](#)
  - Käytännössä tutkijat julkaisevat datansa muualla, valtion tutkimuslaitosten dataa löytyy



**A ilman rekisteröitymistä (CC BY 4.0)**

Sopii tutkimusaineistoihin, joiden jatkokäyttö ei ole rajattua lainsäädännöllisillä tai eettisillä perusteilla.

**B tutkimukseen, opetukseen ja opiskeluun** Ailan rekisteröityneille asiakkaille

**C vain tutkimukseen ja ylempiin opinnäytteisiin** (esimerkiksi väitöstutkimukseen, pro graduun ja ylempään AMK-tutkinnon opinnäytetyöhön) Ailan rekisteröityneille asiakkaille

**D vain erillisellä luvalla** Ailan rekisteröityneille asiakkaille

Täytä alla olevat tiedot.

Luvanantaja ja hänen yhteystietonsa

Ilmoita nimi, sähköposti ja puhelinnumero. Aineiston luovuttaja sitoutuu ilmoittamaan yhteystiedoissa tapahtuvat muutokset kirjallisesti Tietoarkistolle.

Varaluvanantaja, kun luvanantajaa ei tavoiteta

(Esimerkiksi silloin, kun yhteystiedot ovat vanhentuneet tai luvanantaja ei itse voi enää antaa lupaa.)

Ilmoita nimi, sähköposti ja puhelinnumero. Aineiston luovuttaja sitoutuu ilmoittamaan yhteystiedoissa tapahtuvat muutokset kirjallisesti Tietoarkistolle.

# TIETOARKISTO

3 ( 10 )

Versio 2023-05-17/A

Tietoarkisto | Arkistointisopimus

**Jos valitsit kohdan D**, on suotavaa, että tutkimusaineistoa voidaan toimittaa jatkokäyttöön ehdon A, B tai C mukaisesti määräajan jälkeen. Määritä ajankohta, josta alkaen lupaa ei tarvitse enää kysyä.

Tutkimusaineisto on käytettävissä \_\_\_\_ / \_\_\_\_ 20\_\_\_\_ alkaen

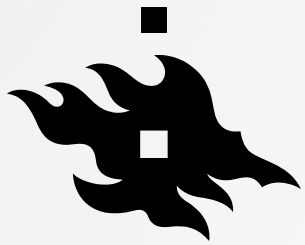
ilman rekisteröitymistä (CC BY 4.0) (ehto A)

tutkimukseen, opetukseen ja opiskeluun Ailan rekisteröityneille asiakkaille (ehto B)

vain tutkimukseen Ailan rekisteröityneille asiakkaille (ehto C).

<https://www.fsd.tuni.fi/files/arkistointisopimus.pdf>





# KIELIPANKKI, END USER LICENSE (EULA)

- CLARIN PUB: loppukäyttäjän lisenssit julkisille aineistoille
  - Sallittu: käyttö, kopiointi, levittäminen, esittäminen, muuntelua, yhdistely
- CLARIN ACA: loppukäyttäjän lisenssit tiedeyhteisön aineistoille
  - Sallittu: käyttö ja kopiointi opiskelu-, opetus- tai tutkimuskäyttöön Käyttäjän oman organisaation sisällä muutettuna, muuttamattomana tai osana yhteenliitettyä teosta
  - Lisäehdot +ID, +BY, +NORED → tunnistautuminen, tekijän nimeäminen, ei jakelua kolmansille osapuolille ilman sopimusta
- CLARIN RES: loppukäyttäjän lisenssit henkilökohtaista käyttöoikeutta edellyttävälle aineistoille
  - Sallittu: käyttö ja kopiointi käyttäjän kanssa sovittuun tarkoitukseen muutettuna, muuttamattomana tai osana yhteenliitettyä teosta.
  - Lisäehdot (+ID, +PLAN, +BY, +NORED) sekä valinnaiset lisäehdot → tunnistautuminen, tutkimussuunitelma, nimeäminen, ei jakelua kolmansille osapuolille ilman sopimusta



# LISENSSIT METADATALLE JA DATALLE

Periaate: **Metatiedot** tulisi avata, vaikka dataa ei voisi avata

- lisenssi yleensä CC0 (metadata = dataa datasta, ei suojaa tekijänoikeus, harvoin tietokantakaan)
- Huom! Ei edelleenkään tutkittavien henkilötietoja, arkaluonteista tai luottamuksellista dataa

**Datalle** “yleismaallinen” ohje: mahdollisimman avoin lisenssi (CC0)

- aina ei kuitenkaan mahdollista (kts. metadatan kohta Huom! ja tekijänoikeus)
- käytännössä ja hyvän tieteellisen tavan mukaisesti aina viitattava alkuperäiseen datan “tekijään”/julkaisijaan (CC BY)
- jos paljon datan julkaisijoita ja paljon datasettejä, CC BY -lisenssin käyttö voi aiheuttaa “viittauskasaantumia” (attribution stacking)



# KIITOS!

- Kysymyksiä, kommentteja?

Bouquet of Flowers, anonymous, 1680. [Public domain](#)

