

Fairdata-verkosto - Tutkimusaineistojen metatietojen haravointi -työpaja 9.6.2021

Aika: Keskiviikko 9.6.2021 klo 13-15

Etäkokoushuone: <https://cscfi.zoom.us/j/67981274279>

Nauhoite: <https://kannu.csc.fi/s/aRBpyoLkrykwQGm>

Wikimerkintä: https://wiki.eduuni.fi/x/_Ai-D

Esitys:



Asialista + keskustelumuistio

1. Tutkimusaineistojen metatietojen haravointi kansainvälisistä palveluista Metaxiin

- Lähtökohtana haravointeihin tutkijan affiliaatio suomalaiseen organisaatioon
- Millaisia toteutuksia tähän mennessä tehty / on suunnitteilla?
 - Jyväskylän yliopisto kommentoinut, että Dryadista on tehty haravointia. Joitakin tietoja on tuotu paikalliseen tietovarastoon, ei vielä tutkimustietojärjestelmään asti. Suunnitelman kanssa ollaan varhaisessa vaiheessa.
- Alustavien käyttötapauksen esittely

2. Metatietojen siirto edelleen kotiorganisaatioiden järjestelmiin

- Nykytilanne Metaxin sisältämien metatietojen hyödyntämisestä
- Muutostarpeet kv. palveluiden haravoinnin näkökulmasta
- Tampereen yliopistosta kommentoitu, että tutkimustietovaranto tulee olemaan työkalu, jolla organisaatioita ja erityisesti yliopistoja mitataan. Kun organisaatioiden on pakko tuottaa ja kaivaa tietoja, työ olisi järkevää toteuttaa yhdessä niin suurelta osalta kuin mahdollista sen sijaan että kaikki käyttävät itse omia vähäisiä resurssejaan. CSC tuskin yksin pystyy tähän, vaan tarvitaan vuorovaikutusta ja yhteistyötä eri organisaatioiden kanssa.
 - Joonas Nikkanen kommentoinut, että aineistojen käyttäminen arvottamiseen tai meritoitumiseen isommin ei ole mahdollista rajallisen kattavuuden osalta. On siis pitkä matka siihen, että aineistoja mitattaisiin metriikan kautta.

3. Metatietojen rinnakkaistallennus – hyvät käytänteet

- Millaisia käytänteitä tällä hetkellä on olemassa?

4. Mentimeter-kysely (kyselyn vastaukset, pdf)

- Onko organisaatiossasi suunniteltu tai toteutettu aineistojen metatietojen haravointia ulkoisista järjestelmistä omiin järjestelmiin?
 - Tampereen yliopistossa suunnitteilla haravointi Tietoarkistosta ensin ja mahdollisesti myös Zenodoon ja EUDATiin edeten, tavoitteena tietojen saanti tutkimustietojärjestelmään. Haravointi/importointi tehdään Metaxin kautta.
 - Aalto-yliopistosta kommentoitu: Puuresta Metaxiin metatietojen integraatio työlistalla. Puuressa (maksuton) integraatio joka haravoi noin 2 000 repositoriota, joskaan ei toimi aivan odotetusti. Repositorioiden metatietokentät eivät ole kovin hyviä, etsii vain affiliaatitietojen perusteella. Metatiedoissa ei ole ollut korjattavaa mutta Pure ei löydä metatietoja niin hyvin kuin mitä todellisuudessa tiedetään olevan. Puuren tulossa, että tutkija voi itse tuoda DOI:n avulla tietoja.
- Tulisiko aineistotietoja kerätä keskitetysti Metaxiin, vai onko suoraviivaisempaa haravoida ne oman organisaation järjestelmään?
 - Miten korjaaminen tehtäisiin ja voisiko organisaatiot olla siinä mukana?
 - Suurin osa keskitetyn ratkaisun puolella, myös haasteita tunnistettu.
- Hyödynnetäänkö organisaatiossasi Metaxin sisältämiä tietoja aineistoista?
 - Jyväskylän yliopistosta kysytty, haravoidaanko tietoa yleisemmin tieteenalakohtaisista repositorioista vai myös geneerisistä repositorioista. Etenkin geneeristen repositorioiden tiedon taso mietityttää.
 - Pohdittu myös, mikä on tutkijan osaaminen tutkijan polulla - esim. roolien erittely voi olla hankalaa kokemattomammille tutkijoille.
 - Aallossa on tullut ilmi, että Zenodossa olevat aineistot ovat moninaisia ja repositorioiden laatuja on monenlaisia. Eroja saako ulosotettavaan tietoon luokiteltua vai tuleeko kaikki tiedot kerralla.
- Minkä arkiston kuvailutietoja toivoisitte haravoitavan Metaxiin ns. pilottitoteutuksena?
 - EUDAT tunnistettu, olisi hyvä että metatiedot siirtyisivät EUDATista Metaxiin ja edelleen Tiedejatutkimus.fi:hin.
 - Fairdatan puolella voidaan kartoittaa tarkemmin pilottitoteutuksia, luultavasti geneerinen repositorio esim. Zenodo tai Dryad – aikataulu aloitus mahdollisesti loppuvuoden 2021 aikana ja 2022.

- Onko aineistotietojen rinnakkaistallentamisesta kertynyt kokemuksia? Millaisia ajatuksia aineistotietojen rinnakkaistallentaminen herättää?
 - [Haravointi on rinnakkaistallentamista mutta on myös tapauksia, joissa itse dataakin saatetaan rinnakkaistallentaa.](#)
 - [Rinnakkaistallentaminen keskeisesti liittyy haravointikysymykseen, tietovirtojen kannalta relevantti asia. Asiasta voitaisiin järjestää oma työpaja.](#)
 - [Palataan asiaan Fairdata-verkoston puolella.](#)
- Muuta palautetta liittyen aineistotietojen haravointiin keskitetysti
 - [Kun tulee uusi tutkija, niin miten pidetään huomiota että hän viittaa oikeaan paikkaan?](#)
 - [Tällä hetkellä pysyvät tunnisteet lähes yksinomaan ratkaisevat ko. ongelman, kun metatietoja siirtyy palvelusta toiseen, pysyvistä tunnisteista linkki alkuperäiseen lähteeseen.](#)
 - [Tutkijan meritoituminen ja vastuullinen arviointi: miten realisoidaan tieto siitä, miten aineistoja käytetään?](#)
 - [Avoimen tieteen seurantomallin luonnos ollut kommentoilla \(<https://docs.google.com/document/d/16ygouUTEhbWAtAaFCxsuqUp3jT7rH93A86dw7LmZbDk/edit>\), kansallista keskustelua aiheesta tarvitaan lisää.](#)

Lopuksi: Työpajan antia käydään läpi Fairdata-verkosto seuraavassa syksyn kokouksessa. Työpajan tallenne, esitys ja kokousmuistio tullaan lähettämään tiedoksi Fairdata-verkoston sähköpostilistan kautta.