

Liisa Näpärä

TUTKIMUSDATAN SÄILYTTÄMISEEN TARVITAAN KANSALLISTA YHTEISTYÖTÄ

Kansalliskirjaston tutkijapalveluiden kehittämistä ja digitaalisten aineistojen käytön edistämistä käyttäjälähtöisellä tiedolla toteuttava Digitaalinen avoin muisti (DAM) -hanke seuraa aktiivisesti digitaalisissa ihmistieteissä käytävää keskustelua. DAM-hankkeen edustajana osallistuin DHN 2020 Norden -etäkonferenssiin. Konferenssissa järjestettiin paneelikeskustelu, jossa pohdittiin digitaalisille aineistojen kohtaloa tutkimushankkeiden jälkeen.

Etäkonferenssista eväitä tutkijapalveluiden kehittämiseen

Riikassa, Latviassa järjestettävä digitaalisten ihmistieteiden pohjoismainen konferenssi (DHN 2020 Norden) siirrettiin maaliskuulta lokakuuhun vain viikkoa ennen kuin se oli aikataulutettu. Näin kävi monelle muullekin konferenssille ja tapahtumalle, kun koko maailma sulkeutui keväällä koronan vuoksi.

Riikan etäkonferenssissa järjestettiin paneelikeskustelu, jossa oli puhumassa James Smithies Lontoon King's Digital Labista, Victoria Johansson Lundin Humlabista ja Jessica Parland-von Essen Suomen CSC:lta. Paneelikeskustelun aiheena oli digitaalisten ihmistieteiden projekteissa syntyneiden datojen ja työkalujen kestävä arkistointi tai säilyttäminen. Kestävä säilyttäminen tarkoittaa sitä, että datat ja työkalut ovat muiden käytettävissä ja toimivat vielä useamman vuoden kuluttua varsinaisen projektin päättymisestä.

Panelistien mukaan merkittävimmät haasteet tällä hetkellä kestävässä datojen ja työkalujen säilyttämisessä ovat projektirahoituksen katkonaisuus, DH-menetelmien nopea kehitys, työkalujen ja formaattien vanhentuminen sekä epäyhtenäiset aineistojen ja työkalujen arkistointikäytännöt. Paneelissa tarjottiin myös vinkkejä, miten haasteita voisi vähentää. Smithies kehottaa tarkastelemaan kokonaisuutta ja arvioimaan säilyttämistä suhteessa haasteisiin sekä panostamaan potentiaalisimpiin datoihin ja työkaluihin. Johansson korostaa tarvetta neuvota tutkijoita käyttämään erilaisia ja muuttuvia datasettejä ja työkaluja, joiden pitkäaikais säilytys riippuu niin ikään niiden ominaisuuksista. Parland-von Essen puolestaan painottaa eri järjestelmien yhteentoimivuutta ja datan löydettävyyden merkityksen tärkeyttä tutkimusprosessin aikana ja sen jälkeen FAIR-periaatteisiin tukeutuen.

Datan ja työkalujen säilyttämistä ei tueta

Keskustelu paneelissa jatkui toteamukseen käytössä olevien dataformaattien epäyhtenäisyydestä ja parhaimpia formaatteja ei osattu oikeastaan edes suositella, vaikka muutamien formaattien todettiin olevan toisia parempia. Panelistit kritisoivat, ettei nykymuotoinen rahoitusjärjestelmä edes tue kovin pitkää datan tai työkalujen säilyttämistä, koska niiden ylläpitoon ei yleensä myönnetä rahoitusta. Paneelikeskustelun loppupuolella

korostettiin yliopistojen ja kansallisen yhteistyön merkitystä, jotta tärkeää dataa ei katoaisi esimerkiksi formaattien muutoksessa tai tulevaisuuden näkökulmasta tärkeän datan merkityksen ymmärryksen puutteen vuoksi.

Kysymykset tutkimusdatan ja tutkimuksessa kehitettyjen työkalujen säilyttämisestä ovat olennaisia myös Kansalliskirjastolle. Perinteisesti sillä ei ole ollut roolia ottaa vastaan tutkimusaineistoa. Tähän mennessä vain pieni osa tutkimuksessa rikastetusta datasta ja työkaluista on voitu ottaa käyttöön sen omissa järjestelmissä, koska niistä on nähty olevan toiminnallisuuksien ja prosessien kehittämisen vuoksi. Ratkaistavaksi jää, voidaanko jatkossa Kansalliskirjaston omiin järjestelmiin integroida nykyistä enemmän tutkimuksissa tuotetuista sisällöistä vai panostetaanko sen sijaan vain sisältöjen löydettävyyteen muista järjestelmistä. Datojen tai työkalujen integrointi ei kuitenkaan ole tarkoituksenmukaista ilman taetta pitkäaikaisesta ylläpidosta ja sisältöjen hyödyllisyydestä suhteessa käyttäjien tarpeisiin. ❖

Kirjoittaja

LIISA NÄPÄRÄ
Kansalliskirjasto
liisa.napara@helsinki.fi

