

# käytäntöön: Tutkimusorganisaation aineistojen säilyttämisestä

Fakta Julkaistu 12.02.2016 klo 15:18 Kirjoittaja Jessica Parland-von Essen

Edellisessä kirjoituksessani kerroin miten tärkeää on keskustella tutkijoiden kanssa heidän työstään, tiedonhallinnastaan ja tarpeistaan kun lähdetään suunnittelemaan tutkimusorganisaation palveluita ja niiden suhteen etenemistä. Toiminnan tavoitteena on tietenkin tutkijoiden työn helpottaminen ja tehostaminen: parempi aineistonhallinta vähentää datan häviämisen, korruptoitumisen tai sensitiivisen datan vuotamisen riskejä. Kaikki nämä voivat olla tutkimukselle hyvinkin kohtalokkaita ja mitä enemmän tutkija on yksin riskienhallinnan kanssa, sitä enemmän siihen myös kuluu aikaa ja energiaa. Hyvä aineistonhallinnan tuki onkin arvokasta työtä, johon kannattaa panostaa koko tutkimusprosessin aikana.



Organisaatio voi edetä monella tavalla tutkimusaineistojen ja tilanteen kartoittamiseen ja aineistonhallinnan haltuunottoon. Tietoisuuden kasvattaminen, neuvonta sekä koulutus ovat ehdottoman tärkeitä toiminnan osia. Lisäksi voi esimerkiksi tarjota tutkijoille helppokäyttöistä palvelintilaa koordinoitusti, jolloin aineistoja saa kerättyä pois omilta kovalevyiltä, muistitikuilta yms. Työkaluja kuten [Cardio](#) tai [DAF](#) voi käyttää tilannekuvan luomiseen, mutta sitä on hyvä tehdä harkiten ja hyvin valmistellen. Tärkeä toimija on usein kirjasto tai tietopalvelu, josta löytyy tiedonhallinnan osaamista ja joka yleensä muutenkin kouluttaa ja palvelee tutkijoita jatkuvasti. Osa tutkijoista saattaa kuitenkin tehdä työtä hyvinkin itsenäisesti, joten etsintääkin kannattaa tehdä.

Kun tutkimusta vielä tehdään aineisto on aktiivista. Se voi jo siinä vaiheessa olla julkista tai sitä ehkä halutaan jakaa tutkijoiden kesken. Silloin se usein pyörii tutkijoiden omissa sovelluksissa ja on epästabiilia. Tutkijan on itse arvioitava mitä tarpeita hänellä on esimerkiksi datan jakamiseen. Olisi hyvä, jos ei tarvitsisi turvautua Dropboxin kaltaisiin kaupallisiin palveluihin. Tosiasia kuitenkin on, että niitä käytetään paljon, sillä ne ovat helppokäyttöisiä ja niissä on jopa tiettyä versionhallintaa. Organisaation palveluilla onkin kilpailuhaastetta, johon vastata. Tutkijan on myös ymmärrettävä vastuunsa: ellei hän ulkoista riskienhallintaa ammattimaisten tutkijapalveluiden tarjoajalle hän joutuu kantamaan sen yksin. Aihe voi kuitenkin olla herkkä ja asian suhteen kannattaakin edetä diplomaattisesti. Tutkijan työtä ei saa kohtuuttomasti hankaloittaa ja hänen tavoitteensa ovat ensisijaisia.

Tutkimusaineistojen elinkaari voi näyttää hyvinkin erilaiselta. Kaikki tutkijat tarvitsevat todennäköisesti tukea ainakin tutkimuksen suunnitteluvaiheessa ja julkaisemisen yhteydessä, jolloin aineistojen on oltava sellaisessa kunnossa, että ainakin muu tutkijayhteisö pystyy arvioimaan tutkimusta. Ensimmäiseen vaiheeseen on tulossa suurta apua uuden kansallisen suunnittelutyökalun [dmpTuulin](#) puolesta, mutta tutkijat tulevat tarvitsemaan tukea sen käyttöön omasta organisaatiostaan.

Entä sitten tuo toinen vaihe, jota voisi kutsua aineistojen luovuttamiseksi? Kun tutkimus julkaistaan, myös datasta tulisi tallettaa kopiot, joihin voi viitata pysyvillä tunnisteilla ja jotka on kuvailtu ja dokumentoitu asianmukaisesti. Joillakin aloilla on tietoarkistoja, jotka voivat palvella tutkijoita, mutta näin ei suinkaan ole kaikilla aloilla. Joka tapauksessa sekä tutkimusorganisaatio että rahoittaja yleensä haluavat tiedot aineistosta ja sen olinpaikasta. Tähän tarvitaan taloudellisia, inhimillisiä sekä teknisiä resursseja ja riittävää osaamista.

Ero tutkimuksenaikaisen aineistojen säilyttämisen ja varsinaisen julkaisemisen tai arkistoinnin välillä on tärkeä hahmottaa. Kun aineisto siirtyy tällaiseen stabiiliin tilaan, vastuut muuttuvat ja on sopimusten teon paikka. Aineiston integriteetti on suojattava; sitä ei enää voi muuttaa. Tutkija tuskin itse haluaa sitoutua ylläpitämään aineiston saatavuutta ja käytettävyyttä pitemmällä aikavälillä. Siksi heti tutkimuksen loputtua on viimeistään sovittava oikeuksista. Osapuoli, jolle aineistot jäävät, joutuu ehkä kopioimaan, konvertoimaan tai migroimaan niitä. Tai ehkä tulevaisuudessa joku muu säilytyspaikka onkin järkevämpi ja ne täytyy siirtää? Tähän on oltava riittävät oikeudet ja kaikkea on sovittava.

Aineistojen pitempiaikaiseen säilyttämiseen liittyy paljon riskienhallintaa, tekniikkaa sekä juridisia vastuita. Siksi tutkimusorganisaationkin kannalta voi olla hyvinkin järkevää ohjata tutkija luotettavaan oman alansa tietoarkistoon. Jos oman alan arkistoa ei löydy, käyttäjille maksuton tutkimusaineistojen säilytyspalvelu IDA voi olla hyvä vaihtoehto. IDA takaa aineistojen turvallisen säilytyksen ja saatavuuden, mutta siihen ei liity tietoarkiston hoivapalveluita.

Tutkimusorganisaation on huolehdittava siitä, että aineisto on kuvailtu riittävän hyvin, jotta sitä voi pitää käyttökelpoisena tulevaisuudessa jos niin halutaan. Käyttöoikeuksista on sovittava selkeästi. Tämä on helpompaa, mikäli tutkijoiden kanssa on jo aiemmin oltu yhteistyössä ja organisaatiossa on toimiva prosessi aineiston ”jäädymisestä” julkaisun yhteydessä ja viimeistään tutkimushankkeen loppuessa. Tutkijoita on syytä opastaa myös eri versioiden tekemiseen tarpeen tullen. Avaamisen hyödyistäkin kannattaa kertoa. Näihin kaikkiin vaiheisiin löytyy palveluita ATT-hankkeessa, ja niiden hyödyntämisestä on syytä sopia tutkimusryhmässä ja hankkeissa.

Mikäli organisaatiossa on riittävää osaamista voi se myös itse ylläpitää tutkimusaineistoarkistoa. Tämä vaatii metatietojen ja pysyvien tunnisteiden hallintaa ja usein käytönhallintaa. [Dspace](#) tai [Fedora](#) voivat ainakin osittain toimia tässä tarkoituksessa, mutta yleisen kuvailun lisäksi on tärkeää saada talteen riittävät alakohtaiset metatiedot sekä huolehtia aineiston löydettävyydestä. Pidempiaikainen säilyttäminen vaatii tarkkaa suunnittelua, paljon metadataa ja riittäviä teknisiä investointeja, minkä takia on hyvä käyttää arvokkaille aineistoille vain sertifioituja tietoarkistoja.

Jos aineiston halutaan olevan käyttökelpoista pitempään, OAIS-viitemallista on apua. Koska aineiston arkistointi on jatkuvaa toimintaa, se edellyttää myös aktiivista hallinnointia ja ylläpitoa. Viitemallin avulla on helpompi saada kiinni siitä, minkälaisia toimintoja sähköiseen arkistointiin karkeasti ottaen kuuluu käsitetasolla. Niihin kaikkiin olisi syytä ottaa kantaa jo kun aineistoja säilytetään, vaikka kaikkea ei itse tarvitse heti tehdä. Esimerkiksi aineiston dokumentaatiota ja kuvailua voi olla mahdotonta tehdä jo muutaman vuoden jälkeen. Jossain vaiheessa kaikkien OAIS-toimintojen tarpeellisuus kuitenkin tulee esiin ja tutkimusorganisaation vastuulla onkin pitemmän aikavälin suunnittelu, johon kuuluvat myös riittävät hallinnolliset rakenteet ja saatavuuden turvaaminen.

OAIS-viitemallista voi lukea lisää [Mari Kleemolan artikkelista “Tietoarkisto toimii OAIS-viitemallin mukaan”](#) (*Tietoarkisto* 2/2010) tai [Brian Lavoien kirjoituksesta “Meeting the challenges of digital preservation: The OAIS reference model”](#)