

Jessica Parland-von Essen

Tutkimusaineistoihin viittaaminen, pysyvät tunnisteet ja linkittäminen

*Jessica Parland-von Essen, orcid.org/0000-0003-4460-3906, Helsingin Yliopisto, CSC,
Jessica.parland-vonessen@helsinki.fi*

Tutkimusaineistoihin viittaamisen käytännöt ja teknologiat ovat keskeisessä roolissa uuden linkitetyn tiedon ja avoimen tieteen maailmassa. Luotettavia yhtenäisiä ratkaisuja tarvitaan sekä tiedon laadun että löytyvyyden vuoksi. Esityksessäni käyn läpi aineistoihin viittaamisen perusteita, pysyvien ja tunnisteiden käyttöä.

Aineistoihin viittaamisen käytännöt eivät vielä ole vielä riittäviä (Mayo et al. 2015, Jonkers et al. 2014, Mayernik et al. 2016), jotta esimerkiksi kyettäisiin luotettavaan viittausten kartoittamiseen tai tekniseen hyödyntämiseen. Vaikka linkittämisen puolella tehdäänkin edistysaskelia (Rueda et al. 2016) se ei asiaa auta, elleivät käytännöt parane. Keskeisissä rooleissa ovat paitsi tutkijat, tietoarkistot ja rahoittajat myös tieteelliset kustantajat ja heidän tuottamansa palvelut. Se millaisia tunnisteita ja viittaussuhteita tarjotaan luo edellytyksiä ja rajoja sille, miten tutkijat voivat niitä käyttää ja miten he omaksuvat niiden tuomat hyödyt.

Suomessa käytetään julkaisu- ja tietoarkistoissa sekä DOI- että URN-tunnisteita. Näillä on sekä hallinnollisia että teknisiä eroja. Kummatkin vaativat aktiivista laadunvarmistusta ja

hallinnointia. Tutkijoita on myös ohjeistettava tunnisteiden käytössä ja viittaamisen käytännöissä.

Tutkimusaineistoihin viittaamisessa pätevät yleensä samat perussäännöt kuin kaikessa viittaamisessa: tavoitteena on varmistaa, että muut tutkijat voivat löytää saman aineiston ja aineiston tekijät saavat heille kuuluvan meriitin. Usein viittaussuhteita löytyy arkistojen verkkosivuilta, mutta myös esimerkiksi kustantajien määräykset vaikuttavat viittaustapaan keskeisellä tavalla. Huolellinen viittaaminen ja tunnisteiden käyttö mahdollistaa myös yhä tärkeemmän konekäytettävyyden, jonka avulla kasvavia tietomääriä on mahdollista hallita ja hyödyntää. Linkittämällä tutkimustuotoksia toisiinsa ne linkittyvät myös tekijöihin, jotka näin saavat näkyvyyttä ja tiedon laatu paranee. (Parland-von Essen 2016)

Hyvät pysyvät tunnisteet ovat hyperlinkkejä, joiden avulla pääsee aineiston luokse verkossa. Pysyvän tunnisteiden voi ilmoittaa vain lyhyenä tunnisteena, mutta ilmoittaminen täydellisenä http-alkuisena verkko-osoitteena helpottaa aineiston luokse pääsyä. Sekä itse aineiston että tunnisteiden elinkaarta on hallinnoitava. Sekä hal-

linnoinnissa ja viittaamisessa on huomioitava esimerkiksi versiot ja aineistojen eri tasot ja niiden osien väliset suhteet. (Kotarski et al 2012).

Käytännön kysymyksiksi muodostuvat siis kysymykset siitä, mitä tunnisteita käytetään, mille tunniste annetaan, mihin tunniste resoluvoituu, miten niiden elinkaarta hallitaan ja lopuksi, muttei vähiten tärkeänä: miten näitä tunnisteita käytetään tutkimuksessa viittausten ja linkitysten yhteydessä? Määrittelyssä tänä vuonna julkaistut FAIR-periaatteet antavat tärkeän ohjenuoran siitä, mihin meidän tulisi pyrkiä. (Wilkinson et al. 2016) FAIR -periaatteet edellyttävät sekä itse aineiston että metadatan hallintaa siten, että esimerkiksi aineiston metatiedot aina ovat saatavilla, vaikka itse aineisto olisi luvanvaraista tai hävitetty. Lisäksi on huolehdittava tietorakenteiden standardinmukaisuudesta ja metadatan laadusta.

Metatietoja on käytännössä eri tasoisia: niitä luodaan yleisemmän löydettävyyden varmistamiseksi kuvailemalla aineisto DataCite tai Dublin Core -formaattien mukaisesti. Silloin on syytä kiinnittää erityisesti huomiota relaatiotietojen riittävyteen ja konekäsiteltävyyteen. Eri versioiden ja aineiston eri osien (esimerkiksi konfiguraatiotiedostojen tai menetelmäkuvausten) suhteiden oikea mallintaminen voi olla haasteellista. Lisäksi aineistoilla on usein tarkempia alakohtaisia metatietoja, joiden formaatti on syytä ilmoittaa. Yhteentoimivuutta kannattaa tukea hyvin kuvailluilla rajapinnoilla. Kontrolloidut sanastot ja ontologiat parantavat tietenkin yhteentoimivuutta.

Metatietojen ja aineistojen eri hierarkkisiin tasoihin tai osiin saattaa syntyä useita eri tunnisteita, joiden hallinnointi vaatii suunnitelmallisuutta. Varsinkin on tärkeä huomioida, että käytettävyys on järjestelmissä suunniteltava siten että loppukäyttäjän on mahdollista hahmottaa selkeä ja yksiselitteinen, kestävä viittaustapa. Useimmat arkistot eivät esimerkiksi tarjoa saman aineiston eri tiedostoille tai rajapinnoille pysyviä tunnisteita, vaan käyttäjälle tarjotaan yhtä tunnistetta, joka resoluvoidaan metatietotietueeseen. Olisi kuitenkin hyvä, jos tarvittaessa kohdesivulta voisi edetä koneellisesti eteenpäin. Myös käyttöoikeuksiin liittyvien metatie-

tojen, esim. pääsynhallinnan, olisi syytä olla koneellisesti suoraan tulkittavissa ja käsiteltävissä.

Tieteelliset kustantajat ovat avainasemassa ohjeistaessaan tutkijoita viittaamisessa.

Lähteet

Jonkers K, Derrick G E, Lopez-Illescas C et al. (2014). Measuring the scientific impact of e-research infrastructures: a citation based approach? *Scientometrics*. doi:10.1007/s11192-014-1411-7

Kotarski R, Reilly S, Schrimpf S, Smit E, Walshe K (2012). Report on best practices for citability of data and on evolving roles in scholarly communication.

Mayernik S M, Hart D L, Keith, E M, Weber, N M (2016). Assessing and tracing the outcomes and impact of research infrastructures. *Journal of the Association for Information Science and Technology*. doi: 10.1002/asi.23721

Mayo C, Hulle E, Vision T (2015). The location of the citation: changing practices in how publications cite original data in the Dryad Digital Repository doi: 10.5281/zenodo.32412

Parland-von Essen J (2016). Tutkimusaineistoon viittaamisen perusainekset, Portti.

Rueda L, Fenner M, Cruse P (2016). DataCite: Lessons Learned on Persistent Identifiers for Research Data, IDCC 2016, practice paper.

Wilkinson Mark D et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* 3/16 doi:10.1038/sdata.2016.18