

Rehtori

13.11.2020

## Lausunto luonnoksesta hallituksen esitykseksi eduskunnalle laiksi tutkimustietovarannosta

Tampereen yliopisto kiittää mahdollisuudesta kommentoida luonnosta hallituksen esitykseksi eduskunnalle laiksi tutkimustietovarannosta.

Lakiesityksen tavoite Suomessa tehtävän tutkimuksen ja tutkimusjärjestelmän julkaistujen metatietojen yhdenmuistamiseksi, ajantasaistamiseksi ja tiedon laadun parantamiseksi sekä tämän kautta saavutettavat hyödyt, avoimen tieteen toteutumisen edistyminen ja sirpaleisen tiedon yhdistämisestä tutkijoille, tutkimusorganisaatioille ja tutkimuksen rahoittajille aiheutuvan lisätyön väheneminen, ovat kannatettavia. Ei ole kuitenkaan kaikilta osin selvää, miten näiden tavoitteiden ja hyötyjen saavuttaminen toteutuu lakiesityksessä esitetyin keinoin.

**Henkilötietojen käsittely.** Lähdejärjestelmissä henkilötietojen käsittelylle on useita käsittelyperusteita (yleinen etu ja suostumus) tutkimustietojärjestelmään siirrettäville tiedoille. Lakiesityksen perusteluissa myös mainitaan, että kukin tutkimustoimija voi itsenäisesti päättää, mitä tietoja tutkimustietovarantoon lähdejärjestelmistä siirretään. Edellä mainitut seikat vaarantavat tutkimustietovarantoon siirrettävien tietojen laajuuden ja yhteismitallisuuden. Näin ollen olisi tarpeen säätää erikseen minimitiedoista, jotka kunkin tutkimustoimijan tai tutkimustoimijaryhmän olisi siirrettävä tutkimustietovarantoon.

Minimitietojen määrittely lakitekstissä voisi osaltaan myös vähentää juridista epäselvyyttä, joka liittyy tietojen siirtämiseen lähdejärjestelmistä tutkimustietojärjestelmään ja siirrettävien tietojen käsittelyyn ja julkisuuteen.

**Tietojen korjaaminen.** Lakiesityksen perusteluissa todetaan toisaalta, että "[t]utkimustietovarannossa tietoja ei muokata, vaan ne ovat aina yksi yhteen lähdejärjestelmien kopioita. Mahdolliset virheiden korjaukset sekä tietojen muuttaminen tai poistaminen tehdään aina lähdejärjestelmässä." (s. 22), ja toisaalta, että "[t]ietojen korjaaminen tarkoittaa tietojen laadun parantamista. Sillä ei tarkoiteta yleisessä tietosuoja-asetuksessa tarkoitettua tietojen korjaamista, mihin rekisterinpitäjä on velvoitettu. Tietojen täydentäminen ja korjaaminen tapahtuu hakemalla tietoja muusta kuin alkuperäisistä lähdejärjestelmistä. Tietojen täydentäminen ja korjaus välitetään myös lähdejärjestelmän tiedoksi." (s. 26). Käytännön näkökulmasta on todettava, että Tampereen yliopiston tutkimustietojärjestelmässä (Pure) ei ole ns. kirjoittavaa rajapintaa, joten tietojen automaattinen siirtäminen tutkimustietovarannosta lähdejärjestelmään kuvatulla tavalla ei ole nykyisellään mahdollinen. Tietojen laadun varmistamisen kannalta on tärkeää, että tietojen muokkaaminen toteutetaan ainoastaan yhdessä paikassa (lähdejärjestelmässä) eikä synny tilannetta, jossa tutkimustietovarannossa ja lähdejärjestelmässä olisi ristiriitaisia tietoja. Käytännön työ on lisäksi osoittanut, että tutkimustietojärjestelmän tietojen automaattinen rikastaminen julkaisijoiden tietokannoista johtaa usein laatuongelmiin ja vaatii manuaalista validointia eli lisätyötä. Jos lakiesityksessä tietojen korjaamisella ei tarkoiteta tietojen korjaamista tietosuoja-asetuksen mielessä, tulisi lakitekstissä käyttää muuta ilmaisua, esimerkiksi täydentäminen.

Rehtori

13.11.2020

**Muut kuin julkaisutiedot.** Tutkimustietojärjestelmään tallennettuja tietoja ei nykyisellään validoida eli laatuvarmisteta kuin julkaisutietojen osalta. Lakiesityksen seurauksena tutkimustietovarantoon on tarkoitus siirtää muitakin tietoja kuin julkaisutietoja. Tämän seurauksena tutkimustietovarannossa on runsaasti tietoja, joiden laatua ja laajuutta ei ole varmistettu. Realistisen kuvan muodostaminen Suomen tutkimusjärjestelmästä tällaisten tietojen perusteella on kyseenalainen. Edellä ehdotettu siirrettävien minimitietojen määrittely poistaisi osittain tämän ongelman, mutta toisaalta edellyttää tutkimustoimijoilta lisäresursseja metatietojen laadun varmistamiseen. Pelkkien julkaisutietojen validoiminen vaatii jo nykyisellään merkittäviä resursseja ja kansallisesti tasalaatuisen tiedon tuottaminen on haastavaa. Tiedonkeruun laajentaminen asteittain muhinkin tutkimustuotoksiin lisää tätä haastetta ja lisäresursoinnin tarvetta entisestään.

**Tutkijatunniste.** Tutkimustietovarantoon siirrettyjen tietojen yhdistely on suunniteltu tapahtuvan tutkijatunnisteen avulla. Lakiesityksessä tutkijatunnisteena esitetään kansainvälinen ORCID-tunniste. ORCID-tunnisteen käyttö on kuitenkin tutkijoille vapaaehtoinen, ja sen käyttöaste vaihtelee tutkimusorganisaatioiden välillä ja sisällä. Näin ollen ilmeisenä vaarana on, että kokonaiskäyttöasteen jäädessä alhaiseksi tietojen yhdistely vaikeutuu merkittävästi, millä on vaikutuksia tutkimustietovarannon käyttöön. Sinänsä ORCID-tunnisteen käyttö on kannatettavaa, ja Tampereen yliopisto toivoo opetus- ja kulttuuriministeriön esittävän keinoja sen käytön lisäämiseen.

**Ohjaussuhteet.** Lakiesitystekstin perusteluista ilmenee, että opetus- ja kulttuuriministeriö käyttää merkittävää ohjausvaltaa suhteessa muihin yhteisrekisterinpitäjiin tutkimustietovarannon käyttöä ja kehittämistä koskien. Varsinaisessa lakiesityksessä tilanne kuitenkin näyttäytyy tasaveroisemmalta yhteisrekisterinpitäjien kesken. Käytännön työn sujuvuuden varmistamiseksi valtasuhteet olisi hyvä selkiyttää jo lakiesityksessä.

Käytännön toteuman näkökulmasta on oleellista, että tutkimustoimijoilla on todelliset mahdollisuudet olla mukana määrittelemässä tarkempia toteumakäytäntöjä ja päätöksenteko on oikea-aikaista siten, että organisaatioille jää tarpeeksi aikaa toteuttaa mahdolliset muutokset (esimerkiksi nykyisellään muutokset opetus- ja kulttuuriministeriön tiedonkeruun käsikirjaan tulevat myöhäisessä vaihteessa ja kesken aktiivisen tiedonkeruukauden).

**Jatkovalmistelu.** Lakiesityksen jatkovalmistelussa tulisi erityisesti huomioida valmistelussa olevan luonnoksen hallituksen esitykseksi laiksi julkisin varoin tuotettujen tutkimusaineistojen uudelleenkäytöstä suhde ja vaikutukset tähän lakiesitykseen.

Erityistä huomiota tulee kiinnittää myös metadatan erittäin keskeiseen rooliin tutkimusaineistojen osalta sekä taustalla vaikuttavaan FAIR-datan olettamukseen. Lain toimeenpano edellyttää niiden myötä tutkimustoimijoiden taholta merkittäviä lisäpanostuksia, jotka toivomme opetus- ja kulttuuriministeriön ottavan huomioon myös tarvittavan resursoinnin näkökulmasta.

Lakiesityksessä esitetään jaettavaksi myös kokonaisia asiakirjoja (tutkimussuunnitelmia, aineistonhallinnan suunnitelmia, tutkimuslupia; vrt. julkisuuslaissa annetut rajoitukset), mikä on merkittävä muutos ja riskitekijä verrattuna aiempiin käytänteisiin. Asiakirjojen sisältöjen luottamuksellisuus ja luottamuksellisen tiedon siirto ovat aivan eri asia kuin keskeisen metatiedon välittäminen ja niiden avoimuus. Tutkijayhteisön luottamuksen menettäminen olisi takaisku varsinaiselle tavoitteelle, ja asiakirjojen sisällöt tulisivat rajata lakiesityksen ulkopuolelle.

Rehtori

13.11.2020

Lakiesityksen taustalla on oletus, että “[k]ansainvälisesti vakiintunut käytäntö on julkaisujen ja tutkimusaineistojen kuvailutietojen avoimuus, vapaa jakaminen, hyödyntäminen ja niihin viittaaminen”. Haluamme tuoda esille, että tämä ei ole vielä käytäntö, vaan ideaali, jonka osa tutkimusyhteisöstä jo omaksunut.

Mari Walls  
rehtori

Valmistelijat Susanna Nykyri, Päällikkö, Avoimen tieteen palvelut, kirjasto  
Turkka Näppilä, datan hallinnan asiantuntija, Avoimen tieteen palvelut, kirjasto  
Petri Tonteri, tietoaaineistopäällikkö, Avoimen tieteen palvelut, kirjasto

Allekirjoitus Asiakirja on sähköisesti allekirjoitettu Tampereen yliopiston asianhallintajärjestelmässä.  
Allekirjoituksen oikeellisuuden voi todentaa Tampereen yliopistolta.