

Kansalliskirjaston lausunto

Viite: Hallituksen esitys laiksi tutkimustietovarannoksi, VN/20733/2020.

Kiitämme mahdollisuudesta kommentoida tutkimustietovarantoa koskevaa lakiesitystä.

Mielestämme tutkimustietovarannolle suunniteltu rooli tutkimustietoa kattavasti kaikkien tutkimuskentän toimijoiden kesken välittävänä palveluna on kannatettava ajatus. Se lisää sekä tutkimuksen että sitä tekevien tutkijoiden näkyvyyttä ja kokoaa samalla arvokasta tietoa tutkimusjärjestelmän ohjauksen tarpeisiin.

Laajassa käytössä olevana julkisena tietovarantona tutkimustietovaranto tasapainottaa tämänhetkistä tilannetta, jossa merkittävä osa tutkimusta koskevasta tiedosta on tallennettu kaupallisten toimijoiden ylläpitämiin tietovarantoihin ja -järjestelmiin. Palvelu mahdollistaa kattavan, luotettavan ja ajantasaisen tutkimusta koskevan tiedon saannin ja lisää tiedon saatavuuteen liittyvää tasa-arvoa sekä tutkijoiden että kansalaisten keskuudessa.

Nähdäksemme lakiesityksessä ehdotettu yhteisrekisterinpitäjäyys OKM:n ja tutkimustoimijoiden kesken on perusteltu ratkaisu. Tietoja on kerätty OKM:n tarpeisiin jo aiemmin, mutta koska tietojen oikeellisuus on tutkimustoimijoiden käsissä, niiden roolia on hyvä nostaa esiin. Tämä lisää myös toiminnan läpinäkyvyyttä ja tuo tutkimuskentän toimijoiden näkemyksiä paremmin mukaan keskusteluun.

CSC:llä on lisäksi keskeinen rooli tutkimustietovarannon ylläpitäjänä ja tietojen käsittelijänä. Monen organisaation yhteistyönä syntyneiden tutkimustuotosten (esim. yhteisjulkaisut) kohdalla niiden raportoimien päällekkäisten tietojen yhdistäminen algoritmien avulla saattaa tosin joissakin tapauksissa mutkistaa tietosisältöjä koskevien vastuiden jakautumista.

Tietojen kokoaminen tietovarantoon edellyttää yhteisten tietomäärittelysten luomista. Niiden avulla tutkimustietovaranto voi välittää yhteismitallista, avointa ja linkitettyä tietoa myös muihin palveluihin. Tämä parantaa tietoaineistojen ja linkitettyä tietoa hyödyntävien tietojärjestelmien yhteentoimivuutta. Yhteisten tietomäärittelysten rinnalla tarvitaan myös pysyviä tunnisteita. Kansainvälisiin standardeihin perustuvien tunnisteiden avulla voidaan taata, että tieto on löydettävissä ja hyödynnettävissä laajasti myös tulevaisuudessa ja sen merkitys säilyy muuttumattomana erilaisissa tietovirroissa.

Näkemyksemme mukaan myös yhteisten tietomäärittelysten avoimuudesta tulisi huolehtia pitkällä tähtäimellä (vrt. esim. digi- ja väestötietoviraston ylläpitämä Yhteentoimivuusalusta).

Tutkimustietovarannon koneluettavat rajapinnat tarjoavat väylän tietojen sujuvaan hyödyntämiseen erilaisissa tutkimukseen liittyvissä palveluissa ja prosesseissa. Tämä vähentää samojen tietojen syöttämistä käsin useaan paikkaan ja säästää siten resursseja. Rajapintojen avoimuus mahdollistaa samalla monenlaisia muitakin käyttötapauksia, sillä tiedot ovat tätä kautta myös hakukoneiden ja tutkimuksen hyödynnettävissä esim. tiedonlouhinnan tarpeisiin. Tutkimustietovarannon tietojen yhdistäminen muihin tietolähteisiin tarjoaa myös runsaasti uusia mahdollisuuksia.

Tältä osin kiinnitimme kuitenkin huomiota siihen, ettei sen paremmin lakiesityksessä itsessään kuin sen perusteluissa oteta selkeää kantaa avointen tietojen jatkokäytön ehtoihin. Lakiesityksen pykälässä 4 todetaan, että "koneluettavasta rajapinnasta tapahtuvalle avointen tietojen lataamiselle ei saa asettaa ehtoja". Vaikka tietovarannon rajapinta olisi vapaasti käytettävissä - ainakin sen teknisen kapasiteetin asettamissa rajoissa - tietojen käyttöoikeuksien selkeä määrittäminen esim. Creative Commons -lisenssien avulla olisi niiden hyödyntämisen kannalta keskeisen tärkeää. Tältä osin toivomme, että tietovarannosta rajapinnan kautta välitettävät avoimet tiedot olisi mahdollista antaa käyttöön CC0-lisenssillä, kuten esim. Kansalliskirjaston ylläpitämässä Finna.fi-palvelussa on jo tehty ja kuten myös eurooppalaisten tutkimusrahoittajien Plan S -aloite tutkimusjulkaisujen kuvailutietojen osalta edellyttää.

Lakiesityksen perusteluissa esitetty jaottelu avointen ja käyttörajattujen tietojen välillä herättää muitakin kysymyksiä. Kuten perusteluissa todetaan, monet tietovarannon sisältämistä tiedoista ovat luonteeltaan julkisia (esim. julkaisujen bibliografiset tiedot, ml. tekijätiedot) ja niitä on perinteisesti levitetty laajasti monenlaisissa yhteyksissä, ml. kirjastojen tietojärjestelmät. Rajanveto avointen ja käyttörajattujen tietojen välillä ei kuitenkaan ole kaikilta osin näin yksinkertaista.

Tutkijaa koskevat tiedot ovat yksi esimerkki alueesta, jolla avoimen ja käyttörajattujen tiedon välinen raja ei ole välttämättä aivan yksiselitteinen. Tutkimustietovarantoon kytketään tutkijan profiilipalvelu, joka antaa My Data -periaatteen mukaisesti tutkijalle itselleen mahdollisuuden hallita omia tietojaan ja niiden käyttöä. Osa tietovarannon sisältämistä henkilötiedoista (esim. julkaisun tekijätiedot) on kuitenkin selkeästi julkisia, ja koska julkaisuja käytetään mm. yliopistojen rahoitusmallissa rahoituksen kriteerinä, niitä koskevien tietojen kattava julkisuus on myös tutkimustietovarannon päämäärien kannalta perusteltua. Tältä osin tutkimustietovarannon toiminta edellyttää selkeää jakoa sen suhteen, mitkä tiedot ovat lähtökohtaisesti julkisia ja mitä tietoja tutkija voi hallinnoida itse.

Henkilötietojen käyttöoikeuksilla on myös merkitystä sen kannalta, onko niitä mahdollista hyödyntää esim. Kansalliskirjaston ylläpitämässä auktoriteettitietokannassa.

Käyttörajattujen tietojen välittäminen järjestelmien välillä edellyttää joka tapauksessa sitä, että tutkimustietovarannon tekniset rajapinnat mahdollistavat käyttörajattujen tietojen rajaamisen vain auktorisoitujen käyttäjien saataville.

Helsinki, 16.11.2020

Liisa Savolainen
Apulaisjohtaja
Kansalliskirjasto

Pekka Heikkinen
Lakimies
Kansalliskirjasto