

EDUSKUNNAN HALLINTOVALIOKUNNALLE

ASIA: HE 74/2021 vp Hallituksen esitys eduskunnalle avoimen datan direktiivin täytäntöönpanoa koskevaksi lainsäädännöksi

Johdanto

Helsingin yliopisto kiittää mahdollisuudesta kommentoida hallituksen esitystä avoimen datan direktiivin kansalliseksi lainsäädännöksi. Yliopiston kaltaiselle avoimia tutkimusaineistoja tuottavalle ja julkisen sektorin tietoja tutkimuskäytössä hyödyntävälle organisaatiolle avoimen datan direktiivi ja sitä täydentävä kansallinen lainsäädäntö on tärkeä. Helsingin yliopisto tukee tiedon avointa saatavuutta, sekä tutkimuksen että koko yhteiskunnan tarpeisiin.

Lakiesityksessä ehdotetulla sääntelyllä on vaikutuksia erityisesti julkisilla varoilla tuotettujen tutkimusaineistojen julkaisemiseen, ja tästä syystä Helsingin yliopiston kommentit keskittyvät tähän osuuteen.

Julkaisemisen määritelmästä ja tavasta sekä lain soveltamisalasta

Helsingin yliopiston tavoitteena on edistää avointa tiedettä ja tutkimusaineistojen avaamista periaatteella ”Niin avointa kuin mahdollista, niin suljettua kuin välttämätöntä”. Tutkimusaineistojen julkaiseminen voidaan tämän periaatteen mukaisesti tehdä monin eri tavoin ja ”eri asteisesti”. Lainsäädännön soveltamisalaksi on määritelty tutkimusaineistojen uudelleenkäytön osalta, että lakia sovelletaan jo julkiseksi saatettuun aineistoon. Tältä kannalta on jäänyt vielä hieman epäselväksi, mitä ”julkiseksi saattamisella” tarkoitetaan. Julkaisemisen jakaminen kahteen kategoriaan ”julkaistu” / ”ei julkaistu” jättää lain soveltamisalan hieman epämääräiseksi käytännön tutkimustoiminnan ja julkaisujen kannalta. Toisaalta on lausuttu, että tätä lakia ei sovelleta, mikäli tiedon avaamista on rajoitettu tietosuojan vuoksi tai tekijänoikeudellisista syistä. Kysymys kuuluukin, jääkö lain soveltamisalan ulkopuolelle, jos on jokin syy rajoittaa aineoston jatkokäyttöä? Onko tämän lain tarkoitus soveltua laisinkaan sellaisiin aineistoihin, joita ei pystytä täysin avoimesti ja vapaasti avaamaan?



UNIVERSITY OF HELSINKI

Aineistot voidaan käytännössä saattaa täysin avoimesti julkiseksi ja vapaasti käytettäväksi tai tietojen käyttöä voidaan kontrolloida eri asteisesti. Osa lain tarkoittamista tietovarastoista voi pitää tutkimusaineistoja täysin avoimena verkossa, toisissa voidaan vaatia kirjautumista ja tunnistautumista ja mahdollisuus on myös vaatia tiedot haluavaa perustelemaan tietopyyntö sekä selventämään tietojen käyttötarkoitus. Tietopyyntöjen perustelu on tarpeen, etenkin jos kyse on henkilötietoa sisältävästä aineistosta tai aineistosta, joka on kerätty suoraan tutkittavilta tai jonka käyttökäyttötarkoitusta on esimerkiksi tietoluvassa rajoitettu.

Henkilötietojen käsite on laaja ja usein eettisistä syistä myöskään anonyymiä aineistoa ei voida jakaa täysin vapaaseen käyttöön (jos esim. tutkimusta tehdään olemalla suoraan kontaktissa tutkittavien kanssa, on kunnioitettava tutkittavan antamaa tietoista suostumusta osallistua tutkimukseen). Mm. tämän vuoksi julkaisua on rajattava erilaisin mekanismein. Helsingin yliopistolle on jäänyt vielä hieman epäselväksi, miten tämä kysymys suhteutuu tämän lain soveltumiseen.

Jos lakia sovelletaan vain sellaisiin aineistoihin, jotka on päätetty saattaa täysin avoimesti saataville ilman kontroleja, niin yliopiston näkökulmasta lain tavoitteet ja velvoitteet toteutuvat pitkälti jo nykyisin.

Arvokkaat tietoaaineistot

Tässä vaiheessa on vaikeaa arvioida konkreettisia vaikutuksia arvokkaiden tietoaaineistojen osalta kun ei ole tietoa komission täytäntöönpanoasetuksesta.

Arvokkaita tietoaaineistoja ovat yliopistolla potentiaalisesti mm. ilmastoa ja maaperää koskeva seurantadata, joka edustaa myös lakiesityksessä kuvattua tiheästi päivittyvää dataa. Arvokkaita tietoaaineistoja voivat olla myös useat lääketieteellisen tutkimuksen aineistot. Lääketieteellisen tutkimuksen lainsäädäntö on myös kansallisella tasolla muutoksessa ja ottaen huomioon tämän lain mahdollinen soveltamisalan rajaaminen, tähän aiheeseen on vaikeaa ottaa kantaa.

Mahdollisia vaikutuksia aineistojen avaamiselle

Helsingin yliopiston näkökulmasta hyvää on se, että laki ei velvoita tutkimusorganisaatiota saattamaan tutkimusaineistoa julkisesti saataville, vaan lakia sovelletaan ainoastaan jo julkaistuun aineistoon. Lisäksi lakiesitys ei edellytä tutkimusaineiston julkaisijalta (tutkija, tutkimusorganisaatio, rahoittaja) ehdotonta veloitetta asettaa tutkimusaineisto saataville lain edellyttämässä muodossa, mikäli tästä olisi kohtuutonta vaivaa, eli toimijoille jätetään esityksessä omaa harkintavaltaa asiassa. Lakiesitys ei myöskään velvoita julkaisijaa muuntamaan jo avattuja aineistoja tai luomaan niistä kokonaan uusia versioita.

Lakiesityksessä todetaan (4.2.3.1., *Vaikutukset tutkimusyhteisöön*), että sääntelyn vaikutukset tutkimusorganisaatioihin jäisivät vähäisiksi. Tämä pitää paikkansa siltä kannalta, että julkisen tutkimusrahoituksen ehtona on pääsääntöisesti tutkimusaineistojen avoin saatavuus jo



UNIVERSITY OF HELSINKI

entuudestaan ja yliopistot ovat sitoutuneet edesauttamaan avointa tiedettä ns. FAIR periaatteita seuraten. Tältä näkökannalta vaikutukset voidaan toki katsoa jäävän pieniksi, mutta väite vaikuttaisi perustuvan puutteelliseen tietämykseen siitä, millainen prosessi tutkimusaineiston julkaiseminen on, ja mitä se edellyttää sekä tutkijalta että tutkimusorganisaatiolta. Lisäksi OKM:n syksyllä 2019 tekemän avoimuuden kypsyystasoselvityksen yhteydessä suoritettiin tutkimusorganisaatioissa kysely keskeisistä avoimen tieteen esteistä. Jo ennestään suurimmaksi esteeksi tutkijat ovat ilmoittaneet epätietoisuuden siitä, täyttääkö aineistojen avaaminen juridiset vaatimukset: epävarmuus ja monimutkaisuus voi johtaa siihen, että aineistoja ei avata.

Tutkimusaineistojen uudelleenkäyttö vaatii suunnitelmallisuutta aineiston keräämisestä alkaen. Lainsäädännöllä ei tulisi lisätä tutkijoiden taakkaa ja viedä aikaa varsinaiselta tutkimustyöltä. Käytännössä kaikki lisävelvoitteet kuitenkin merkitsevät lisää työtä tutkijoille, joten direktiivin soveltamisen mukanaan tuovien velvoitteiden hoitamiseen on osoitettava tarvittavat resurssit tutkijoille ja tutkimuksen tukipalveluille. Kustannuksia ja vaatimuksia aiheutuu niin yliopistolle kuin tutkijallekin. Työtaakan kasvu voi myös vähentää tutkijan halua asettaa tutkimusaineistoja julkisesti saataville.

Direktiivi (yhdistettynä muuhun tutkimukseen liittyvään lainsäädäntöön) voi kääntyä alkuperäistä tarkoitustaan vastaan, jos se aiheuttaa sen, että aineistoja ei sitten avatakaan koska lakia ei haluta sovellettavan niihin tai se tuntuu liian vaikealta.

Mahdollinen taannehtiva soveltaminen

Lainsäädännön on ilmeisesti tarkoitus tulla sovellettavaksi silloin kun aineisto on *jo* avattu. Lakiesityksessä ei kuitenkaan oteta selkeästi kantaa siihen mahdollisuuteen, että jo ennen lain voimaantuloa avattu tutkimusaineisto ei täyttäisi lain vaatimuksia. Lain soveltaminen takautuvasti jo julkaistuihin aineistoihin ei olisi tarkoituksenmukaista ja oletamme että tämä ei ole tarkoitus.

Esimerkki: v. 2019–2020 siirrettiin kaksi Helsingin yliopiston isoa tutkimusaineistoa, joista toinen sisälsi *jo avoimesti saatavilla olevaa* SMEAR-dataa, digitaaliseen pitkäaikaissäilytykseen. Aineistojen kuratointi kymmeniä vuosia kestävää pitkäaikaissäilytystä varten edellytti viikkojen työn sekä tutkijoilta että tukipalveluiden edustajilta.

Lisäksi kaikki havaintolaitteet eivät toistaiseksi ole niin automatisoituja, että niiden tuottama data olisi reaaliaikaisesti avoimen rajapinnan kautta saatavilla. Sääntelystä voi aiheutua huomattavia kustannuksia yliopiston tutkimusinfrastruktuureille, missä tapauksessa on tärkeää, että arvokkaiden tietoa-aineistojen maksuttomuudelle tulee direktiivin kuvaama siirtymäaika (2 vuotta).

Kustannusvaikutukset

Lakiesityksessä mainitaan ainoastaan eräille julkisen sektorin elimille (esim. kirjastot, arkistot, museot) lain myötä syntyvät tulonmenetykset aineistojen maksuttomasta käytöstä, mutta ei käsitellä lainkaan sitä, että myös tutkimusorganisaatioille syntyy merkittäviä välillisiä kustannuksia tutkimusaineistojen avaamisesta – sekä nykyisellään että lakiesityksen voimaantulon jälkeen.

Tutkimusorganisaatiot ovat sekä tutkimusaineistojen pääasiallisia tuottajia että kouluttavat aineistojen analysoinnin asiantuntijat. Tutkimusaineistojen uudelleenkäyttö edellyttää myös investointeja: aineistojen avaaminen ei ole sitä, että aineistot vaan laitetaan verkkoon vaan aineistojen avaamiselle tarvitaan palvelut ja henkilöstöresurssit. Kustannukset on huomioitava tutkimusaineistoja tuottavien ja rahoittavien organisaatioiden rahoituksessa. Lisäksi on syytä ottaa huomioon mahdollisesti myös yksityisten tutkimusrahoittajien (esimerkiksi säätiöt) rahoittamissa hankkeissa syntyvät aineistot. Niiden uudelleenkäyttöä ei välttämättä aina ole myönnettyssä rahoituksessa otettu huomioon, jolloin kustannukset tulevat tutkimusorganisaatiolle.

Julkinen vai yksityinen rahoitus?

Lakiesitys koskee vain julkisella rahoituksella tuotettuja tutkimusaineistoja, joten se jättää avoimeksi sen, millaisia velvoitteita tutkijoilla tai yliopistoilla on yksityisellä rahoituksella tuotettujen avoimien aineistojen suhteen. Tutkimusaineistoja tuotetaan Suomessa myös esimerkiksi yksityisten säätiöiden rahoituksen tuella. Miten lakia sovelletaan niissä tapauksissa? Onko tarkoitus, että yliopistolla julkisoikeudellisena laitoksena on velvollisuus huolehtia myös siitä, että nämä aineistot täyttävät lakiesityksen mukaiset edellytykset, kun yliopisto julkisena organisaationa tukee (ja rahoittaa) tutkimusta tarjoamalla palvelut ja tutkimusinfrastruktuurin tutkimuksen käyttöön? Muuttaako tämä tutkimusaineiston julkisesti rahoitetuksi?

Reaaliaikaisesti päivittyvän tiedon julkiseksi saattamisesta

Helsingin yliopisto tuottaa mm. ilmastoa ja maaperää koskevaa seurantadataa, joka edustaa lakiesityksessä kuvattua tiheästi päivittyvää tietoa sekä mahdollisesti myös arvokkaita tietoaaineistoja. Lakiesityksessä julkisen hallinnon tiedonhallinnasta annetun lain muuttamisesta todetaan, että tiedon tulisi olla pyynnöstä saatavilla teknisten rajapintojen avulla heti tiedon keräämisen jälkeen.

Lakiesityksen muotoilu ”heti” jättää jonkin verran avoimeksi sen, missä rajoissa tutkimuksen aikainen datan kerääminen kuuluu lain piiriin. Tiedon keräämiseen sisältyy useita prosesseja: havainnon rekisteröinti mittalaitteella, tiedon prosessointi ja syöttäminen tietokantaan jne. Asiaa selventäisi esim. muotoilu ”viipymättä tiedon keräämisen jälkeen”. Toisaalta direktiivissä ja



kansallisessa laissa annetaan joustoa aikataulun suhteen, jos velvollisuuden noudattaminen aiheuttaa viranomaiselle kohtuutonta vaivaa taloudellisesti ja teknisesti, mikä toki antaa jo hieman liikkumavaraa.

Tutkimusaineiston avaamisen kannalta keskeisiä kysymyksiä

Yleisenä lähtökohtana on jo ennestään, että aineistot ovat mahdollisimman avoimia ja saavutettavia FAIR-periaatteiden (Findable, Accessible, Interoperable, Reusable) mukaisesti.

Aineistohallinnan suunnittelu

Tutkimusaineistojen organisoitu avaaminen edellyttää tutkijalta ja organisaatiolta aineistohallinnan suunnittelun osaamista, jotta varmistetaan, että aineistojen metatiedot on oikein laadittu ja lainsäädännölliset velvoitteet esim. tietosuojan ja tietoturvan osalta ovat kunnossa. Aineistojen koko elinkaari, mukaan lukien uudelleenkäyttömahdollisuus, on huomioitava jo ennen aineiston keräämistä aineistohallintasuunnitelmassa.

Tutkimusaineistojen uudelleenkäytettävyys

Lakiesityksessä esiintuotu uudelleenkäytettävyys ei ole avatun tutkimusaineiston automaattinen ominaisuus: mikäli sitä ei ole otettu huomioon aineistohallinnan suunnittelussa, tutkimusaineiston kokoamisvaiheessa ja aineiston metatietoja laadittaessa. Tutkimusaineiston uudelleenkäytettävyydessä ei ole kysymys pelkästään siitä, että avatulla aineistolla on jatkokäytön mahdollistava lisenssi, vaan myös siitä, että aineiston kuvailutiedot on riittävän tarkasti laadittu. Uudelleen käytettävyyden varmistaminen edellyttää tutkijoilta tutkimusaineistojen avaamisen osaamista ja erillistä työpanosta sekä yliopistolta tukipalveluita, koulutusta ja perehdytystä.

Lakiesityksen vaarana on, että jo avatun tai avattavan tutkimusaineiston uudelleenkäytettävyyden varmistamisesta voi muodostua yliopistolle merkittäviä lisäkuluja. Vaikutusten arvioinnissa olisi voinut vielä käsitellä tutkimusaineistojen uudelleenkäytön edellyttämiä investointeja sekä näiden investointien vaikutuksia kansantaloudelle ja yhteiskunnalle.

Kaupallisesta ja yleisen edun mukaisesta käytöstä

Lakiesityksessä on mainittu, että tutkimusaineistoja voidaan avata kaupalliseen tai muuhun kuin kaupalliseen käyttöön. Jää epäselväksi, kuka tai mikä taho tämän viime kädessä määrittää, ja missä määrin julkaisijalla on lain voimaantulon jälkeen oikeus vaikuttaa tähän. Onko valinta kaupallisen tai muun kuin kaupallisen käytön välillä julkaisijan päätettävissä? Vai onko lakiesityksen sisältönä se, että kun aineisto on avattu, se on annettava vapaasti sekä kaupalliseen että tutkimuskäyttöön? Mitä siinä tapauksessa tapahtuu niille avoimeksi saatetuille aineistoille, joiden kaupallista käyttöä on rajoitettu, tai joiden käyttöä rajaa tietosuojalainsäädäntö tai joihin tutkija on asettanut muita ehtoja? Jäävätkö ne kokonaisuudessaan lain soveltamisalan ulkopuolelle?

Aineistojen uudelleen käytettävyyttä saattaa rajoittaa se, että aineisto on kerätty tai luovutettu yliopistolle yleisen edun mukaiseen tieteelliseen tutkimukseen, tällöin aineistojen avaaminen



UNIVERSITY OF HELSINKI

avoimesti ilman rajoituksia myös kaupalliseen tarkoitukseen tulee ongelmalliseksi. Tutkijoilla ja tutkimusorganisaatioilla tulisi säilyä valta päättää, mihin käyttöön aineistoja avataan. Tämä liittyy usein myös arvoihin sekä tutkijan vapauteen ja arvomaailmaan. Tutkijalla voi olla perusteltuja syitä rajoittaa aineiston käyttömahdollisuuksia. Kaupallisesta käytöstä tehdään normaalisti erillinen sopimus yrityksen kanssa, joka tutkimusaineistoa haluaa käyttää ja tämän tulisi olla mahdollista myös jatkossa.

Helsingissä 27.5.2021

Paula Eerola
Tutkimuksesta vastaava vararehtori

Lausunnon valmistelijat:

Tiina Käkelä
Asiantuntija

Lotta Ylä-Sulkava
Lakimies, HY:n tietosuojavastaava