

Asia: VN/16660/2020

Lausuntopyyntö luonnoksesta hallituksen esitykseksi avoimen datan direktiivin täytäntöönpanoa koskevasta lainsäädännöstä

Tiedonhallintalain ja julkisuuslain muutokset

Onko ehdotuksella vaikutuksia, joita ei ole vielä tunnistettu, erityisesti lakien soveltamisalassa oleviin organisaatioihin? Millaisista vaikutuksista on kyse tai miten niitä olisi mahdollista arvioida tarkemmin?

Ehdotuksessa ja direktiivissä ei velvoiteta rajapintojen, jotka olisivat nimenomaan avoimia, käyttöön. Mitä mieltä olette väitteestä, että viranomaiset oman harkintansa mukaan käyttäisivät avoimia rajapintoja joka tapauksessa silloin, kun se on tarkoituksenmukaista?

-

Onko ehdotettujen muutosten johdosta tarpeen muuttaa myös jotain erityislainsäädäntöä, jota ei esitysluonnoksessa ole tunnistettu?

-

Laki eräiden julkisten yritysten tietojen uudelleenkäytöstä

Kommentit laista eräiden julkisten yritysten tietojen uudelleenkäytöstä kokonaisuutena.

-

Millaisiin yrityksiin toimialallanne tai alueellanne ehdotettu laki tulisi sovellettavaksi?

-

Tunnistatteko julkisia yrityksiä koskevaa sääntelyä, jota olisi ehdotetun lain johdosta muutettava tai joka on sen kanssa ristiriidassa?

-

Onko tiedossanne sääntelyä, joka velvoittaa ehdotetun lain soveltamisalassa olevia julkisia yrityksiä antamaan niiden hallussa olevia asiakirjoja uudelleenkäytettäväksi?

-

Onko tiedossanne sääntelyä, joka velvoittaa ehdotetun lain soveltamisalassa olevia julkisia yrityksiä julkaisemaan tai julkistamaan niiden hallussa olevia asiakirjoja, mutta jonka tosiasiallisena tarkoituksena tai vaikutuksena on tietojen uudelleenkäyttö kaupallisiin tai muihin tarkoituksiin?

-

Onko ehdotetun lain soveltamisalassa oleviin yrityksiin tai muihin toimijoihin sellaisia vaikutuksia, joita ei ole tunnistettu? Millaisista vaikutuksista on kyse tai miten niitä olisi mahdollista arvioida tarkemmin?

-

Onko asiakirjojen uudelleenkäytön salliminen ominaista tai olennaista jonkin ehdotetun lain soveltamisalassa olevan julkisen yrityksen liiketoiminnalle?

-

Laki julkisin varoin tuotettujen tutkimusaineistojen uudelleenkäytöstä

Kommentit laista julkisin varoin tuotettujen tutkimusaineistojen uudelleenkäytöstä kokonaisuutena.

Suomen tieteellinen kirjastoseura pitää lakiehdotusta tervetulleena, vaikkain tavoitteisiin pääsyn varmistamisessa on vielä runsaasaasti täsmennettävää. Avoimen datan direktiivin (2019/1024) implementointi ja datan uudelleenkäyttöä koskeva lainsäädäntö vie eteenpäin niitä tavoitteita, joita kansallisella ja kansainvälisellä avoimen tieteen yhteistyöllä edistetään. Tutkimusprojekteille tehtävästä aineistohallintasuunnitelmasta tulee entistäkin tarpeellisempi dokumentti, ja avoimien lisenssien käytölle datan julkaisemisessa on nyt selkeämmät perusteet.

Lain soveltamisala jää kuitenkin epäselväksi. Laki koskisi julkaistavia tutkimusaineistoja, mutta varaumia aineistojen julkaisemiselle on jätetty. Juridisten esteiden lisäksi datan avaamista ei tarvitsisi tehdä, jos ei olisi tarkoituksenmukaista ja vaiva on kohtuuton (4 §). Jos näihin tullaan vetoamaan, lainsäädännön tavoittelemat hyödyt voivat jäädä vähäisiksi. Rahoittajien vaatimukset ja suositukset olisivat edelleen tärkein kannustin tutkimusaineistojen avaamisessa.

Onko ehdotuksella vaikutuksia tutkimusaineistoja tuottaviin tai julkaiseviin tahoihin, joita ei vielä ole tunnistettu? Millaisista vaikutuksista on kyse tai miten niitä olisi mahdollista arvioida tarkemmin?

Tutkimukseen liittyvällä aineistohallinnan suunnittelulla on entistä suurempi merkitys ja tämä pitäisi ottaa huomioon tutkimusorganisaatioissa ja kansallisissa tutkimustiedon hallinnan ratkaisuisissa. Suomessa on sitouduttu direktiivissä ja lakiehdotuksessa mainittujen FAIR-periaatteiden toteuttamiseen. Datan tulisi olla löydettävää (Findable), saavutettavaa (Accessible), yhteentoimivaa (Interoperable) ja uudelleenkäytettävää (Re-usable). Kansallisia datanhallinnan palveluita kehitetään näiden pohjalta. Käytännössä FAIR-periaatteiden mukaan toimiminen vaatii todella paljon työtä kaikilla tasoilla ja kaikilta toimijoilta, jotka osallistuvat datan hallintaan sen elinkaaren aikana.

Aineistohallintasuunnitelmalla (data management plan, DMP) osoitetaan, miten FAIR-periaatteita tutkimusprojektissa käytännössä toteutetaan. Aineistohallintasuunnitelmassa tunnistetaan juridisia ja eettisiä riskejä, joita avaamisella voi olla. Jos näitä riskejä tai esteitä on, dataa ei avata. Datan

kuratointiin sekä tallennusratkaisuihin tutkimuksen aikana ja tutkimuksen jälkeen on kiinnitettävä entistä enemmän huomiota. Aineistoja tallennetaan kansainvälisiin data-arkistoihin (repositorioihin), jotka voivat olla tieteenalakohtaisia tai monille eri aloille suunnattuja. Data-arkistojen taso on vaihtelevaa, ja sopivan tallennuspaikan valinnassa tutkijat tarvitsevat usein apua. Suositus on, että aineiston metadata avattaisiin, vaikka dataa kokonaisuudessaan ei pystyttäisi avaamaan.

Yhteistyössä tuotetusta tutkimusdatasta tulee eri toimijoiden kesken laatia sopimukset, joissa aineistojen avaamisesta sovitaan. Tämä käytäntö pitää saada rutiiniksi, koska vastuiden määrittelystä ja niiden hoitamisesta tulee entistä tärkeämpi osa tutkimusprosessia. Tässä korostuu neuvontapalveluiden saatavuus eli tutkimuksen tukipalvelujen tarve tulee lisääntymään.

Datanhallinnan osaajia on tutkimusorganisaatioissa vielä vähän eikä suoraa koulutusta alalle ole. Tieteelliset kirjastot ovat avainasemassa datanhallinnan käytäntöjen kouluttamisessa, ja tähän tarvitaan entistä enemmän henkilöstöä. Laajempi osaajapula on tunnistettu myös Euroopan komissio datastrategiassa (2020), jonka mukaan EU-alueella on 1 miljoonan data-ammattilaisen vaje.

Onko olemassa riskiä siitä, että tutkimusaineistoja ei enää asetettaisi julkisesti saataville sen vuoksi, että ehdotettu laki tulisi niihin julkaisun johdosta sovellettavaksi?

On. Tieteenalojen omat käytännöt vaikuttavat datanhallintaan ja avaamisen ratkaisuihin, ja tässä on todella paljon vaihtelua. Avatun datan tulisi olla mahdollisimman laadukasta, jotta se olisi myös käytettävää. Esimerkiksi laitteiden mittausdataa ei välttämättä saada suoraan käytettävässä muodossa. Koneellisesti luettavalla muodolla tarkoitetaan tiedostomuotoa, jonka rakenne mahdollistaa määrättyjen tietojen yksilöinnin, tunnistamisen ja poimimisen, mukaanlukien tietoalkiot ja niiden rakenne. Tutkimusdata on kuitenkin usein rakenteetonta ja vaatii rakentaistamista. Datan ja metadatan kuratointi vaatii resursseja, kaikki ei vielä tapahdu automaattisesti. Datan avaamiseen ja säilyttämiseen liittyviin kustannuksiin ei myöskään osata varautua. Jos henkilöitä, aikaa tai rahaa ei ole, julkaisemista ei tapahdu.

Missä laajuudessa tutkimusaineistojen maksutonta saatavuutta edellytetään jo nykyään osana julkista tutkimusrahoitusta?

Suomen Akatemia edellyttää tutkimusaineistojen avointa saatavuutta. Perustelluista syistä aineistolla voi olla erilaisia avoimuuden asteita. Suunta on sama suurten kansainvälisten rahoittajien osalta. Rahoituksen saaminen edellyttää organisaation hyväksymän aineistohallintasuunnitelman, jossa mm. eritellään mahdollinen aiemmin kerätty data ja sen käyttö, miten ja missä tutkimusprojektin data tallennetaan, avataan ja tutkimuksen tulokset julkaistaan. Jos dataa tai tutkimuksen tuloksia ei voida julkaista avoimesti, siihen on oltavat perustelut.

Tunnistattekko tutkimusaineistoja, niiden tuottamista tai hyödyntämistä koskevaa sääntelyä, jota olisi ehdotetun lain johdosta muutettava tai joka on ehdotetun sääntelyn kanssa ristiriidassa?

Lakiehdotuksessa käsitelty sääntely (tietosuoja, tekijänoikeus ja lähioikeudet, teollisoikeudet, liikesalaisuudet, julkisuuslaki) ovat niitä osa-alueita, joita datanhallinnan suunnittelussa yleisesti käsitellään.

Arvokkaat tietoaineistot

Onko toimialallanne tai toiminnassanne tunnistettu tietoaineistoja, jotka mahdollisesti tai todennäköisesti tultaisiin komission toimesta säätämään arvokkaiksi tietoaineistoiksi?

Tieteellisillä kirjastoilla ei ole arvokkaita tietoaineistoja, mutta kehysorganisaatioiden tutkimustoiminta voi tuottaa niitä, jolloin kirjastojen datapalvelut voivat osallistua niiden avaamiseen.

Onko toimialallanne tai toiminnassanne varauduttu arvokkaiden tietoaineistojen sääntelystä johtuviin vaikutuksiin? Miten kuvailisitte vaikutuksia ja niihin varautumista? Antakaa mahdollisuuksien mukaan myös euro- tai lukumääräarvioita.

-

Yleiset kommentit

Yleiset kommentit hallituksen esityksen luonnoksesta ja ehdotetun sääntelyn vaikutuksista kokonaisuudessaan.

Lainsäädännön myötä korostuu datanhallinnan suunnittelun tärkeys ja osaaminen. Siinä määritellään julkaisemisen tapoja ja tuetaan tutkimusaineistojen hyödyntämistä myös kaupalliseen käyttöön. Joillekin tutkimusaloille tämä on jo arkipäivää, mutta ei kaikille.

Lakiesityksessä ja direktiivin tekstistä saa vaikutelman, että tiedeyhteisön sisällä on olemassa valmiit käytännöt, joiden toteuttaminen olisi rutinoitunutta. Tilanne ei kuitenkaan ole tämä. Esimerkiksi avoimen datan direktiivin yhteydessä käyty keskustelu metatietojen minimimallista on täysin kesken.

Avoimen datan direktiivin suomenkielinen käännös sisältää muutamia termejä, jotka ovat siirtyneet osittain myös lakitekstiin ja voivat aiheuttaa väärinymmärryksiä. Direktiivin johdanto-osan kappaleessa 27 englanninkielisen version termi 'data management planning' on käännetty harhaanjohtavasti 'tiedonhallinnasuunniteluksi'. Se on tiedonhallintalain mukainen käsite. 'Data management planning' on suomeksi aineistonhallinnan (tai datanhallinnan) suunnittelua. Aineistonhallintasuunnitelma (data management plan, DMP) on tutkimusprosessin avaindokumentti, jota useat rahoittajat ja entistä useammat tutkimusorganisaatiot edellyttävät tutkimusprojekteilta. Lakiehdotuksessa on myös käytetty 'data repository' termistä käännöstä 'tietovarasto'. Tutkimuksen osalta yleisemmin käytetään termiä data-arkisto tai repositorio. Tietovarasto (data warehouse) on

tekniseltä käytettävyydeltään erilainen ratkaisu eikä sitä esiinny direktiivin englanninkielisessä versiossa.

Kaiponen Päivi
Suomen tieteellinen kirjastoseura - Päivi Kaiponen, puheenjohtaja