

Lausunto

23.11.2020

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?

Rajapintojen tulee olla ensisijaisesti tarkoituksenmukaisia. Esimerkiksi pieniä, staattisia aineistoja voi olla mielekkäämpää jakaa tiedostomuodossa (esimerkiksi .csv-taulukko tai .xml-tiedosto) kuin rajapinnan kautta. Rajapintojen avoimuus helpottaa datan uudelleenkäyttöä. Avoimet rajapinnat ovat myös viranomaiselle käytännöllinen tapa tarjota dataa, koska datan uusiokäyttäjän ei tarvitse tilata haluamaansa aineistoa erikseen, vaan hän voi ladata sen itse. Näin ollen halutun aineiston toimittaminen ei vaadi erillistä työtä viranomaiselta.

Rajapintojen avoimuus, standardointi ja yhdenmukaisuus helpottaa myös erilaisten ja eri lähteistä peräisin olevien datojen yhdistelyä ja uudelleenkäyttöä. Vaikkei rajapintojen avoimuutta erikseen vaadita, niiden vapaaehtoinen, harkintaan perustuva käyttö ja siihen kannustaminen on kannatettava tavoite.

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

Juridiset esteet datan liikkuvuudelle on identifioitava ja poistettava, ja datan liikkuvuutta on edistettävä horisontaalisesti eri sektoreiden välillä, jotta datan potentiaali pystytään mahdollisimman hyvin hyödyntämään. Esimerkiksi EU:n tekijänoikeusdirektiivi ei kaikilta osin tue datatalouden tavoitteita täysimääräisesti. Sen kansallisessa soveltamisessa tulee pyrkiä

hyödyntämään koko kansallinen liikkumavara datatalouden periaatteiden kanssa yhteensopivaksi. Esimerkiksi datan uudelleenkäyttö tiedon- ja tekstinlouhinnan menetelmin tulisi olla mahdollista nykyistä laajemmin.

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.

CSC näkee hyvänä, että julkisilla varoilla tuotettujen aineistojen avoimuutta ja sitä kautta jatkokäyttöä edistetään järkevillä tavoilla. Lähtökohtana tulee olla, että aineistot ovat mahdollisimman avoimia ja saavutettavia FAIR-periaatteiden mukaisesti, silloin kun se on mahdollista, huomioiden lainsäädäntö, etenkin yksityisyydensuojan näkökulmasta.

Aineistojen koko elinkaari, mukaan lukien uudelleenkäyttömahdollisuus, on huomioitava jo ennen aineiston keräämistä aineistohallintasuunnitelmassa. Monet tutkimusorganisaatiot ja -rahoittajat vaativat aineistohallintasuunnitelmia jo nyt, ja tätä kehitystä on tuettava. Aineistot eivät yleensä

ole jatkokäyttökelpoisia sellaisenaan, vaan aineiston sisällön, keruutavan ja muuttujien kuvailu on välttämätöntä aineiston käyttökelpoisuudelle. Tätä kuvailevaa tietoa kutsutaan metatiedoksi. Hyvät metatiedot myös helpottavat tutkimusaineiston löytämistä. Metatietojen kattavuutta ja yhdenmukaisuutta on varauduttava tukemaan suositusten, ei velvoittavan säätelyn pohjalta.

Myös aineistosta, joka ei ole kerätty nimenomaan tutkimustarkoitusta varten, voi tulla tutkimusaineistoa. Aineisto voi muodostua myös useammasta eri lähteestä kootuista tiedoista tai aineistoista, jonka osana voi olla eri syistä luottamuksellista aineistoa, esimerkiksi liikesalaisuuksia koskien. Näiden osalta direktiiviluonnos voi aiheuttaa rajanvetotapauksia.

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

Tutkimusaineistojen avoimuudella edistetään tutkimusaineistojen jatkokäyttöä, josta hyötyvät aineiston tuottajan itsensä lisäksi muut tutkijat ja parhaimmillaan yhteiskunnat maailmanlaajuisesti. Tutkijat voivat etsiä käyttöönsä tarvitsemaansa dataa ja tutkijat hyötyvät itsekkin oman datansa avaamisesta. Avoimuus parantaa tutkijan näkyvyyttä: avoimella tausta-aineistolla varustetut julkaisut keräävät enemmän viittauksia kuin julkaisut, joiden tausta-aineisto ei ole saatavilla. Itse datakin voi tuoda tutkijalle viittauksia.

Avoin data parantaa tutkimuksen läpinäkyvyyttä ja toistettavuutta, ja tätä kautta tieteen laatua. Avoimuus myös demokratisoi tiedettä: tutkijoilla vähävaraisemmista yliopistoista ja kolmannen maailman yliopistoista on yhtäläinen pääsy tutkimuksen tuotoksiin kuin varakkaiden yliopistojen tutkijoilla.

Yhteiskunnan suurten haasteiden ratkaiseminen edellyttää poikkitieteellistä, dataintensiivistä tutkimusta, jossa eri alojen dataa on voitava globaalisti yhdistellä, eli datan saatavuus ja uudelleenkäyttävyyys valtion- ja tieteenalarajojen yli on yhtä tärkeämpää.

Tutkimusaineistojen uudelleenkäyttö vaatii suunnitelmallisuutta aineiston keräämisestä alkaen. Se ei kuitenkaan saa lisätä tutkijoiden taakkaa ja syödä näin aikaa varsinaiselta tutkimustyöltä. Käytännössä kaikki lisäveloitteet merkitsevät lisää työtä tutkijoille, joten direktiivin soveltamisen mukanaan tuovien veloitteiden hoitamiseen on osoitettava tarvittavat resurssit tutkijoiden työkaluille ja tutkimuksen tukipalveluille. Tutkimusorganisaatioissa on viime vuosina lisätty tiedon avoimuuden toteuttamiseen vaadittavaa tukea ja koulutusta, ja tätä on jatkettava. On myös huomioitava, että osa direktiivin veloitteista vaatii teknisiä ratkaisuja ja sovelluskehitystä, mikä on huomioitava organisaatioiden rahoituksessa.

Aineistojen säilyttäminen ja uudelleenkäytön mahdollistaminen aiheuttaa kustannuksia. Lisäksi aineistojen käytettävyydestä on tarpeen mukaan huolehdittava. Tällaisia kuratointitarpeita voi

syntyä etenkin teknologian kehittyessä, tallennusmenetelmien ja aineistojen käyttöön tarvittavien ohjelmistojen päivittyessä. Jotkut aineistot voivat lisäksi vaatia tukea uudelleenkäyttöön.

Tutkimusaineistojen uudelleenkäytön mahdollistamisen kustannukset on arvioitu olevan noin 5% tutkimusrahoituksesta. Direktiivin toimeenpanossa tämä tulee huomioida tutkimusaineistoja tuottavien ja rahoittavien organisaatioiden rahoituksessa (esimerkiksi korkeakoulut, tutkimuslaitokset, Suomen Akatemia, Business Finland).

Tutkimusaineistojen uudelleenkäyttöä lisätään velvoittavan sääntelyn ohella myös luomalla tutkijoille ja tutkimusorganisaatioille insentiivejä ja kannustimia. Näitä tulee kehittää yhteistyössä tutkimusyhteisöjen ja -rahoittajien kanssa. Kansainvälinen yhteistyö on tärkeää, jotta varmistetaan datan yhteentoimivuus eri tasoilla.

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

CSC pitää mahdollisena, että pitkäaikainen säilytys- ja saatavuusvelvoite saattaa joissakin tapauksissa johtaa tilanteeseen, jossa aineistoja poistetaan saatavilta juuri ennen lain voimaantuloa, jotta välttyttäisiin tulevilta velvoitteilta näiden aineistojen osalta.

Direktiivin soveltaminen ei saa johtaa kohtuuttomiin tilanteisiin esimerkiksi edellä kuvattujen kuratointitarpeiden vuoksi. Datan avaaminen ja julkaiseminen tulisi olla tutkijoille ja tutkimusorganisaatioille aina hyödyllistä ja kannustettavaa myös niiden omasta näkökulmasta. Tähän tarvitaan sääntelyn ohella kulttuurisia ja taloudellisia kannusteita. Tutkijayhteisöjä tulee tukea siinä, että datan avaaminen ja uudelleenkäyttö muodostuvat nykyistä vahvemmin akateemisen meritoitumisen muodoiksi.

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

Kansallisella tasolla tutkimuksen avoimuutta, mukaan lukien tutkimusaineistot, on voimakkaasti lisätty yhteisillä linjauksilla ja periaatteilla. Tässä työssä ovat olleet mukana niin tutkijat, korkeakoulut, tutkimuslaitokset, muut tutkimusorganisaatiot, tutkimusrahoittajat sekä ministeriö. Tällä hetkellä merkittävimmät rahoittajat edellyttävät avoimuutta mahdollisuuksien mukaan. Myös Euroopan komission Horizon 2020 -puiteohjelmassa aineistojen avoimuutta pääsääntöisesti vaaditaan.

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

-

Arvokkaat tietoaaineistot

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

-

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

Keskeisenä tieteen ja tutkimuksen toimijana ja osana suomalaista tutkimusinfrastruktuuria, sekä kansallisten pitkäaikaissäilytysratkaisujen tuottajana CSC:llä on valmius vastata tuleviin vaatimuksiin palvelutarjonnassaan.

Yleiset kommentit

Yleiset kommentit hallituksen esityksen luonnoksesta ja ehdotetun sääntelyn vaikutuksista kokonaisuudessaan.

CSC - Tieteen tietotekniikan keskus Oy (CSC) kiittää mahdollisuudesta lausua luonnoksesta hallituksen esitykseksi avoimen datan direktiivin täytäntöönpanoa koskevasta lainsäädännöstä. CSC on Suomen valtion ja korkeakoulujen omistama erityistehtäväyhtiö. Tarjoamme teknologian ja palvelukehityksen ratkaisuja tutkimuksen, koulutuksen, kulttuurin ja julkishallinnon erityisosaamisalueilla. Ratkaisujemme perustana toimivat ICT-alustat, valtakunnallinen tutkimus- ja koulutusverkko Funet sekä tehokkaat konesalitoiminnot. CSC:llä on merkittävä rooli tiede- ja koulutuspoliittisen ohjauksen ja kehittämisen välineenä sekä kulttuuripolitiikan toimeenpanon tukena.

Tiede hyötyy datan avoimesta jakamisesta, uudelleenkäytöstä ja analysoinnista maailmanlaajuisesti eri tieteenaloilla. Suuret, kansalliset rajat ylittävät haasteet, kuten ilmastonmuutokseen liittyvät ongelmat, on mahdollista ratkaista vain tutkimuksen poikkitieteellisellä ja avoimella yhteistyöllä. Avoin pääsy dataan edistää myös kansalaisten osallistumista tutkimusprosesseihin ja parantaa tutkimuksen avoimuutta sekä poliittisen päätöksenteon läpinäkyvyyttä ja demokratiaa. Elinkeinotoiminnan kytkeminen vahvemmin avoimen tieteen hyödyntämiseen on tärkeää.

CSC kannattaa tiedon ja tieteen avaamista FAIR-periaatteita noudattaen. Julkisin varoin tuotetun tiedon avaaminen on tärkeää demokratian ja tasa-arvoisuuden näkökulmasta. Lisäksi avoimuus tukee julkisia organisaatioita koskevaa hyvän hallinnon periaatetta. Tämä kuitenkin vaatii datan yhteentoimivuuden varmistamista kaikilla tasoilla ja tarvittavien lainsäädäntömuutosten tekemistä, osaamispääoman systemaattista kehittämistä sekä hallinnonalojen välisten siilojen purkamista.

Datamäärien räjähdysmäinen kasvu ja big datan hyödyntäminen vaativat tieteellistä laskentaa, jotta data pystytään jatkojalostamaan sellaiseen muotoon, joka on yhteiskunnan ja elinkeinoelämän hyödynnettävissä esimerkiksi työpaikkojen ja yritystoiminnan luomisessa. Datan liikkumista rajojen, tieteenalojen ja sektoreiden välillä tulee aktiivisesti edistää, ja datanhallintaan ja käsittelyyn tarkoitettuja infrastruktuureja tulee määrätietoisesti kehittää.

Datan liikkuvuuden ja käytettävyyden takaamiseksi Suomessa tulee huomioida sekä kansallinen että kansainvälinen yhteentoimivuus. Suomen tulee edistää omassa lainsäädännössään ja kansainvälisessä toimintaympäristössään yhteisiä datapolitiikkoja, kuten FAIR-periaatteita. Avoimen datan direktiivin kansallinen soveltaminen annettua liikkumavaraa maksimaalisesti hyödyntäen, ylimääräistä sääntelyä välttämällä, on erinomainen lähtökohta.

Yleisesti datatalouden kehittämisen pitää tarkoittaa, että tuetaan aidosti datapohjaisia palveluita kehittäviä startup-kasvuyrityksiä ja luodaan niiden liiketoiminnalle suotuisat ja houkuttelevat olosuhteet. Eri sektoreiden – kaupallinen, tutkimus, julkinen – välistä datan liikkuvuutta, yhdistelyä, uudelleenkäytettävyyttä ja löydettävyyttä tulee aktiivisesti tukea. Tässä on tärkeää soveltaa edellä mainittua yhteentoimivuuden viitekehystä kaikilla sen tasoilla, lainsäädännöllisten esteiden kartoittamisesta ja purkamisesta teknisiin sovelluksiin.

Tieteen ja liike-elämän yhteinen kehitystyö vaatii datan liikkuvuutta eri sektoreiden välillä, ja siksi on erityisen tärkeää Suomen tasolla purkaa hallinnonalojen väliset siilot, jotta päästään aitoon dialogiin ja yhteistyöhön. Uusia datainfrastruktuureja luotaessa on tärkeää rakentaa olemassa olevien rakenteiden päälle, jotta vältetään päällekkäisyys ja voidaan taata yhteentoimivuus ja jatkuvuus. Jatkuvuuden takaamiseksi datainfrastruktuureille tulee myös kehittää kestäviä rahoitusmalleja.

Datan avoimuus ja yhteentoimivuus vaativat lisäksi, että datan käsittelyyn on olemassa sekä teknisiä resursseja että osaamista. Suurteholaskentaa, datanhallintaa, tekoälyä ja yhteyksiä tulee kehittää kokonaisuutena ja rakentaa synergisiä ekosysteemejä. Dataosaaminen on tärkeää niin Suomen kuin EU:n kilpailukyvyille, ja dataosaamisen kehittämiseen kaikilla sektoreilla ja koulutuksen tasoilla on panostettava.

Toimiva datainfrastrukturi on edellytys myös tekoälyn kehittämiselle. Suomella on mahdollisuus toimia edelläkävijänä tekoälyn kehittämisessä, mutta tekoälyä ei voi eristää omaksi sektorikseen, vaan se vaatii järjestelmällistä panostamista tehokkaaseen laskentainfrastruktuuriin, datan vapaaseen liikkuvuuteen ja osaamisen kehittämiseen.

Hyppölä Jenni
CSC-Tieteen tietotekniikan keskus Oy