

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?

Ehdotuksen vaikutuksia lakien soveltamisalassa olevien organisaatioiden hankintamenettelyä tulisi arvioida, kun julkisin varoin hankitaan palveluita, joiden tuloksena syntyy arvokasta dataa, esim. julkisen sektorin tuottamien tekstien käännöksiä ja julkisen sektorin toimintaa esitteleviä valokuvia.

JYSE PALVELUT ehdot eivät huomioi datan avaamista. Julkisen sektorin tarjous- ja sopimus käytäntöihin tulee luoda avoimen datan hankintaehdot, jotka voidaan antaa esim. tiedonhallintalautakunnan suosituksen osana. Julkisin varoin hankitun datan uloslisensiointi avoimena datana ei onnistu, ellei tällaista dataa lisensioida sisään datan avaamisen mahdollistavilla ehdoilla.

Jos julkisen sektorin hankkima data halutaan avata avoimena datana, tulee JYSE PALVELUT <http://vm.fi/documents/10623/307565/JYSE+palvelut/564be8e8-8ed1-4b44a206-701e54f6348e> ehtoja käytettäessä julkisen sektorin toimijan tarjouspyynnössä ja sopimuksessa määritellä, että oikeudet saadaan julkisen sektorin toimijan käyttöön JYSE ehdoista poikkeavalla tavalla, avoimen datan julkaisemisen mahdollistavassa laajuudessa. Tällöin oikeudet tulee ostaa Creative Commons Nimeä Kansainvälinen Julkinen lisenssi 4.0 tai CC0 luopumaehdoilla avoimen datan julkaisemisen mahdollistamiseksi.

Kun yliopistolta ostetaan palveluja esim. opintoprojekteja tai tutkimusprojekteja, jotka toteutetaan julkiselle sektorille, tulee harkita, onko esim. ministeriöllä tarve saada tuloksiin yksinoikeus, vai voitaisiinko tulosten avoimen julkaisemisen avulla saavuttaa yhteiskunnallinen vaikuttavuus ja opiskelijoiden, yliopistojen ja tutkijoiden jatkotutkimusoikeudet ja start-up- sekä muiden yritysten jatkokäyttöoikeudet. Jos avoimen datan julkaiseminen valitaan, tulee sopia, että projektin tulokset

julkaistaan Creative Commons Nimeä Kansainvälinen julkinen lisenssi 4.0 ehtojen mukaisesti. Näin on menetelty esim. Aalto-yliopiston Design for Government projekteissa.

Uudelleenkäytön oikeudet Suomen julkishallinnon avoimiin tietoineistoihin suositellaan JHS suosituksessa myönnettäviksi avoimella käyttöluvalla. Suositeltavat käyttöluvut ovat on Creative Commons Nimeä 4.0. ja Creative Commons CC0 luopuma. Suosituksen tarkoituksena on tehdä suosituksen kohteena olevan Creative Commons luvan ja luopuman sisältö ymmärrettävämmäksi julkishallinnon käytännön soveltamistilanteissa <http://www.jhs-suositukset.fi/suomi/jhs189>

JYSE ehdot on laadittu siitä näkökulmasta, että ne ovat yleisehdot, joista osaavien hankintayksiköiden pitää osata poiketa, jos halutaan saavuttaa datan avoimen julkaisemisen tavoite.

Hankintayksikön tulee jo ennen tarjouskilpailua määritellä onko tuloksena sellaista materiaalia, jolle tulisi taata vapaa oikeus jatkokäyttöön avoimen datan periaatteiden mukaisesti. JYSE ehtoja ei siis muuteta, vaan sovitaan laajemmat käyttöehdot, jotka kirjataan tarjouspyyntöön ja sopimukseen. JYSE ehtojen tarjoamien oikeuksien lisäksi tilaajan tulee saada oikeudet materiaaliin siten, että julkisen sektorin toimija voi lisensoida materiaalit esim. Creative Commons Nimeä 4.0 Kansainvälinen Julkinen Lisenssi ehtojen mukaisesti <http://creativecommons.org/licenses/by/4.0/legalcode.fi>

Jos avoimen datan vaatimukset huomioidaan hankintojen tarjouspyynnöissä ja hankintasopimuksissa voidaan materiaaleja käyttää esim. European Language Resource Coordinationin tarvitsemin tavoin <http://www.lr-coordination.eu/>

Kieliaineistoja voitaisiin käyttää mm. seuraavasti: Collect Language Resources that can be used to train and/or assess Language Technology tools. Collect tools that can be used to process information from various sources, including translation, extraction or support to fake news detection.

Julkisen sektorin hankkimat tekstit, joissa sama sisältö on käännetty usealle kielelle, ovat erinomaisia materiaaleja esimerkiksi käännöksiä tuottavan tekoälyn kouluttamiseen, jos ne hankitaan avoimen datan julkaisemisen mahdollistavilla ehdoilla. JYSE Palvelut -ehdot vahvistettiin 2.7.2014. JHS 189 suositus annettiin 11.12. 2014. JHS suositusta ei ole huomioitu JYSE ehdoissa.

Avoimen datan julkaisemisen mahdollistava valinnainen lauseke julkisen sektorin tarjouspyyntöön ja hankintasopimukseen tulee julkaista esim. osana tiedonhallintalautakunnan suosituksia.

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?

Teknisten rajapintojen käyttö on hyödyllistä koska silloin data on ajantasaista ja rajapintojen käyttö vähentää datan käsittelyyn vaadittava työmäärää. Ihmisten tekemä työ ja datan kokoaminen eri lähteistä lisää datan virheitä, jos rajapintoja ei voida käyttää. Se mitä dataa jaetaan, tulisi suunnitella pitkäjänteisesti ja datan tuottajien yhteistyön tarpeista lähtien. Tarpeellisuus ja hyödyllisyys arvioinnin avulla löydetään avointen rajapintojen käytön tilanteet. Kriittisten tietojen osalta (kuten VIPUNEN-tiedot) velvoittavuus datan jakamiseen avointen rajapintojen avulla voi olla perusteltua.

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.

Hallituksen esityksen luvussa 4.2.1 jossa kuvaillaan ehdotusten taloudellisia vaikutuksia, olisi tullut käsitellä kansatalouden kannalta keskeistä tutkimusdatan uudelleenkäytön edellyttämää investointitarvetta ja näiden investointien avulla saavutettavia tutkimusaineistojen uudelleenkäytön merkittäviä myönteisiä taloudellisia vaikutuksia koko yhteiskunnassa. Vaikka lakiesitys velvoittaa vain rajallisiin toimenpiteisiin, on tutkimusdatan uudelleenkäyttö datatalouden keskeinen tekijä.

Yliopistoilla on datan tuottajana ja datan käsittelyyn ja analysointiin kouluttavana organisaationa osaaminen toimia datatalouden veturina, mutta datan uudelleenkäyttö vaatii lisäresursseja.

Tutkimusdatan uudelleenkäyttö edellyttää investointeja, joiden suurus on n. 5 % tutkimushankkeiden kustannuksista, näin arvioi raportissaan Commission High Level Expert Group on the European Open Science Cloud sivulla 17 "We expect that on average about 5% of research expenditure should be spent on properly managing and stewarding data used (page 17): https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf ja Bared Mons 25.2.2020 Nature -lehden artikkelissa: "Invest 5% of research funds in ensuring data are reusable" <https://www.nature.com/articles/d41586-020-00505-7>

Tutkimusdatan on oltava löydettävissä, saavutettavissa, yhteentoimivaa ja uudelleenkäytettävissä, Findable, Accessible, Interoperable, Reusable, eli FAIR dataa. Vain FAIR periaatteiden mukaan käsiteltyä dataa voidaan uudelleenkäyttää ja vain FAIR periaatteiden mukaan käsitelty data voidaan avata avoimena datana.

EU:n komissio suosittelee huomioimaan yliopistojen keskeisen roolin datatalouden ekosysteemissä: Cost-benefit analysis for FAIR research data, Recommendation 36: Place universities at the heart of the European FAIR data community of practice. Universities are the most important type of research

performing organisations and are the ones also preparing the workforce of tomorrow who will be able to support the implementation of FAIR data principles in Europe.

Lähde: Cost-benefit analysis for FAIR research data <https://op.europa.eu/en/publication-detail/-/publication/d3766478-1a09-11e9-8d04-01aa75ed71a1>

Datan uudelleenkäytön kustannukset on huomioitava yliopistojen rahoituksessa. Tämä rahoitus on investointi Suomen datatalouteen. EU tasolla tutkimusdatan uudelleenkäytön laiminlyönti aiheuttaa noin n. 26 miljardin taloudelliset menetyksen vuosittain, EU komission selvitys Cost of not having FAIR research data : "Following this approach, we found that the annual cost of not having FAIR research data costs the European economy at least €10.2bn every year. In addition, we also listed a number of consequences from not having FAIR which could not be reliably estimated, such as an impact on research quality, economic turnover, or machine readability of research data. By drawing a rough parallel with the European open data economy, we concluded that these unquantified elements could account for another €16bn annually on top of what we estimated. These results relied on a combination of desk research, interviews with the subject matter experts and our most conservative assumptions.

Lähde: Cost of not having FAIR research data https://www.ouvrirlascience.fr/wp-content/uploads/2019/03/Cost-Benefit-analysis-for-FAIR-research-data_KI0219023ENN_en.pdf

Lakiehdotuksen 6 § on äärimmäisen vaikeaselkoinen. Säännöksen perustelujen esimerkki on kirjoitettu tavalla, joka vaikuttaa sekoittavan lakiesityksen tarkoittaman sallivan avoimen lisensioinnin ja kaupallisen käytön kieltävät copyleft lisenssit. Kun tutkimusdata on julkaistu direktiivin edellyttämällä lisenssillä, joka sallii kaupallisen käytön ja on koneluettava, on tutkimusdata tällöin kaikkien toimijoiden käytössä myös kaupallisesti. Creative Commons CC BY 4.0 julkaistu data on myös julkaisevan yliopiston käytettävissä esim. tilaustutkimushankkeessa.

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.

-

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

-

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

Toimintamalli, jossa data julkaistaan ensin ei-kaupallisen käytön sallivalla lisenssillä ja julkaistun datan yhteydessä on tieto siitä, mistä voi ostaa kaupallisen käytön mahdollistavan lisenssin, ei ole enää mahdollinen. Datan kaupallinen hyödyntäminen tulee toteuttaa kaupallistamisen mallilla, jossa datan saa haltuunsa vain korvausta vastaan.

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

Maksutonta saatavuutta ei edellytetä, sillä tutkimusrahoituksen edellytyksenä olevan yhteiskunnallisen vaikuttavuuden tutkimuksen tuloksille voi saavuttaa sekä kaupallistamalla että avoimesti julkaisemalla.

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

Hallituksen esityksestä ei ilmene, millainen vastuu lain noudattamisesta on tutkijan työnantajalla. Jos yliopiston tutkija on tehnyt teoskynnyksen ylittävän tietokannan osan itsenäisesti, ei yliopistolla ole mahdollisuutta huolehtia tutkimusaineiston uudelleenkäytöstä, koska julkisesti rahoitettu tutkimusaineisto on tekijänoikeuslain 40 b §:n 2 momentin tarkoittamalla tavalla tutkijan yksinoikeuden piirissä.

Tekijänoikeuslain 40 b §:n 2 momentti vaikeuttaa tutkimusaineistojen uudelleenkäyttöä, sillä tietokantoja luodaan usean tutkijan toimesta usein pitkän ajanjakson kuluessa. Tietokantojen osalta on vaikea selvittää, millä rahoituksella mikäkin osa tutkimusdatasta on tehty, onko tietokanta sui

generis tietokantasuojan piirissä vaiko teoskynnyksen ylittävä teos ja onko tietyn tutkijan panos tutkimusdatan tietyn osan tuottamiseen määriteltävissä itsenäiseksi vaiko epäitsenäiseksi työksi.

Arvokkaat tietoaineistot

Onko toimialallanne tai toiminnassanne tunnistettu tietoaineistoja, jotka mahdollisesti tai todennäköisesti tultaisiin 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.

-

Yleiset kommentit

Yleiset kommentit hallituksen esityksen luonnoksesta ja ehdotetun sääntelyn vaikutuksista kokonaisuudessaan.

-

Saaristo Antti
Aalto-yliopisto