

Webinaari 20.11.2020: Plan S:n vaatimukset julkaisuarkistoille ja tutkimustietojärjestelmille

Aika: Perjantai 20.11.2020 klo 9-11

Paikka: Zoom-etäkokous

Eurooppalaisten tutkimusrahoittajien Coalition S -yhteenliittymän lanseeraama [Plan S](#) pyrkii nopeuttamaan siirtymistä tieteellisten julkaisujen välittömään avoimuuteen. Plan S kohdistaa vaatimuksia myös julkaisujen rinnakkaistallentamiseen käytettävälle julkaisuarkistoille.

Kansalliskirjaston yhdessä Suomen Akatemian ja Finn-ARMA-verkoston julkaisutiedonkeruuryhmän kanssa järjestämä webinaari tarjoaa perustiedot Plan S:n vaatimuksista ja mahdollisuuden niitä koskevaan keskusteluun. Tilaisuus on suunnattu julkaisuarkistojen ja tutkimustietojärjestelmien parissa työskenteleville ja muillekin aihepiiristä kiinnostuneille.

Tilaisuus on ilmainen, ja siihen voi ilmoittautua keskiviikkoon 18.11. asti [tällä verkkolomakkeella](#). Etäkokouksen osoite lähetetään ilmoittautuneille tilaisuutta edeltävänä päivänä.

Tervetuloa mukaan!

Tilaisuudesta antaa lisätietoja Jyrki Ilva, [jyrki.ilva\[at\]helsinki.fi](mailto:jyrki.ilva[at]helsinki.fi).

Huom! Suomen Akatemia järjestää pe 27.11. iltapäivällä yleisemmän julkaisujen avointa saatavuutta ja tutkijoiden vastuullista arviointia käsittelevän webinaarin (ks. [ohjelma](#), ilmoittautuminen päättyy 20.11.).

Seminaarin ohjelma 20.11.

Alka	Aihe	Kuvaus	Esiintyjät
9.00-9.05	Tilaisuuden avaus		Jyrki Ilva, Kansalliskirjasto
9.05-9.25	Plan S:n esittely	Plan S ja sen julkaisuarkistoille kohdistamat vaatimukset ja suositukset (pptx , pdf) Ks. myös Suomen Akatemian tiedote ja Jyrki Hakapään blogi-kirjoitus	Jyrki Hakapää, Suomen Akatemia
9.25-9.35	Yliopiston näkökulma	Plan S:n vaatimukset: Näkökulma Turun yliopistosta (pptx , pdf) Taulukko Plan S -kriteereistä julkaisuarkiston arviointiin (pptx, pdf)	Mikko Pennanen, Turun yliopisto
9.35-10.15	Keskustelua Plan S:n teknisistä vaatimuksista ja suosituksista	Keskustelun pohjana käytettävät kalvot (pdf) - Plan S:n makrotason vaatimukset - Julkaisuarkistoa koskevat yleiset vaatimukset (OpenDOAR-rekisteröinti, tuki pysyville tunnisteille, vakaa saatavuus, toimiva palveluosoite) - Täytämmekö nämä jo? - Mikrotason vaatimuksiin liittyviä kysymyksiä - Metadatan CC0-lisenssi - Tietojen tallentaminen koneluettavassa muodossa metadatan lisäksi myös itse dokumenttiin? - Metadatan laadukkuus (artikkelin tunnisteet, lisenssi, versio, OA-status, tutkimuksen rahoittaja) - Vaikutukset tallennusprosesseihin? Mitä tietoja olisi mahdollista saada esim. Tutkimustietovarannosta? - Plan S:n ohjeistuksessa on pakollisten vaatimusten lisäksi myös vahvoja suosituksia. Miten niihin pitäisi suhtautua?	
10.15-10.45	Muita rinnakkaistallentamista koskevia kysymyksiä	- Julkaisun oikeuksiin (<i>rights retention -strategiaan</i>) liittyvät käytännöt ja ohjeistukset: Miten kirjoittajan pitää toimia submittoidessaan julkaisuaan kustantajalle? - Lisenssit: Plan S edellyttää tiettyjen CC-lisenssien käyttöä myös rinnakkaistallentamisessa. Miten tämän kanssa kannattaa toimia? - Rinnakkaistallentamisen aikataulu: Plan S edellyttää rinnakkaistallennukselta välitöntä avoimuutta ilman embargoa - mutta miten nopeasti avoimuuden pitää käytännössä toteutua?	
10.45-11.00		Loppukeskustelu: miten tästä eteenpäin?	

Tilaisuuden muistio

Seminaarin järjestämisen taustalla on julkaisuarkistojen asiantuntijaryhmässä havaittu tarve Plan S -tiedon jakamiseen ja siihen liittyvään keskusteluun. Tilaisuutta ei tallenneta, mutta kalvot tulevat saataville, ja tilaisuudesta tehdään yhteenveto.

Suomen Akatemia järjestää oman tilaisuutensa aiheesta viikon kuluttua perjantaina 27. marraskuuta, ja siihen voi ilmoittautua vielä tänään.

Plan S:n esittely, Jyrki Hakapää, johtava tiedeasiantuntija, Suomen Akatemia

Jyrki H. kertoi työskentelevänsä avoimen tieteen kysymysten parissa ja olevansa Akatemian edustajana mm. Coalition S -verkoston kokouksissa. Jyrki H. piti [yleisesityksen](#) Plan S -vaatimuksista.

Aikataulusta kävi ilmi, että Plan S otetaan käyttöön ensi vuoden alusta eli 1.1.2021 alkaen, ja se koskee sen jälkeen avautuvia Akatemian hakuja. Tämä tarkoittaa, että käynnissä olevien ja tänä vuonna avattavien hankkeiden julkaisut eivät ole sen piirissä, vaan niitä koskevat vielä vanhat toimintatavat. Näin ollen kultainen ja vihreä reitti sekä hybridimalli ovat niiden osalta mahdollisia, samoin embargot. Ehdot tulevat sovellettaviksi vasta niissä Akatemian rahoittamissa hankkeissa, joiden rahoitushaut avautuvat vuodenvaihteen jälkeen, eli ne koskevat mm. huhtikuussa -21 avautuviksi suunniteltuja ohjelmahakuja. Julkaisuarkistojen kohdalla Plan S alkaa siten olla käytännössä ajankohtainen vuoden 2021 lopulla tai -22 alkupuolella. Lisäksi rights retention -strategian soveltamisen osalta odotellaan vielä lisätietoja valmisteilla olevasta uudesta tekijänoikeuslaista.

Rahoituksesta Jyrki kertoi, että jatkossa avoimuuden aikaansaamiseen liittyvät kulut siirretään välillisiksi kuluiksi, kukin suorituspaikka ottaa ne huomioon omassa yleiskustannuskertoimessaan. Myös julkaisuarkistojen ylläpitokulut kuuluvat näihin kuluihin.

Esityksen jälkeen keskusteltiin tekijänoikeuslakiin tulevista muutoksista ja niiden vaikutuksista Plan S -linjauksiin. Jyrki H. totesi, että lain valmistelu on opetus- ja kulttuuriministeriön työtä. Välttämättä avoimuuden voi joku tapauksessa nähdä sopimuksena tutkijan ja rahoittajan välillä, eikä kustantajan ole mahdollista tehdä päällekkäistä, toisensisältöistä sopimusta tutkijan kanssa. Plan S on siten toteutettavissa nykyisenkin tekijänoikeuslain puitteissa, mutta koska lain tiedetään muuttuvan keväällä, on järkevää odottaa ja katsoa, millaisia muutokset ovat.

Yliopiston näkökulma, Mikko Pennanen, tietoasiantuntija, Turun yliopiston kirjasto

[Esityksessään](#) Mikko jaotteli Plan S:n vaatimukset makrotason eli julkaisuarkistolle asetettuihin vaatimuksiin sekä mikrotason eli julkaisukohtaisiin vaatimuksiin. Nämä jakautuivat edelleen pakollisiin ja suositeltuihin kriteereihin. Pakolliset vaatimukset Mikko oli koennut [erilliseen taulukkoon](#). Mikko nosti esiin sen, että rahoittajaa koskevan metadatatiedon selvittäminen sekä metadatan lisääminen varsinaiseen tiedostoon on työlästä. Metadatan tiedostoon saamiseksi yksi ratkaisu voisi olla sen sisältämän kansilehden generointi julkaisuarkistossa. Lisäksi Mikko toi esiin sen, että OKM:n julkaisutiedonkeruussa OA-statuksen raportointiin käytettävät kentät ovat muuttumassa, millä voi olla merkitystä myös Plan S:n vaatimien tietojen osalta.

Keskustelua ja kysymyksiä

Keskustelun tukena käytettiin Jyrki Ilvan tekemää [kalvoesitystä](#). Kalvoissa viitattiin mm. vuonna 2017 valmistuneeseen [Metadatasuositukseseen julkaisuarkistojen tekstiaineistoille](#). Keskustelua herättivät etenkin mikrotason vaatimukset, jotka kohdistuvat julkaisuihin.

Pakolliset vaatimukset

Makrotason vaatimuksista todettiin, ettei niiden toteuttamiseen liity erityisiä ongelmia.

Mikrotason vaatimukset käytiin läpi yksi kerrallaan:

- Standardien mukaista, CCO-lisensioitua metadattaa vaatii myös mm. [Finna.fi](#)-palvelu. Julkaisuarkistojen metadatan lisenssiä ei kuitenkaan ole kaikkien arkistojen osalta määritelty selkeästi, eli tämä vaatii vielä tarkentamista. Metadatan lisenssiä koskevaa tietoa ei ehkä ole tarpeen merkitä erikseen yksittäisiin tietueisiin, vaan todennäköisesti riittää, että se on määritelty arkiston tasolla. Julkaisujen abstraktit ylittävät teoskynnyksen ja ovat tekijänoikeuden alaista aineistoa, eli ellei asiasta ole sovittu, niille ei voi ainakaan automaattisesti antaa CCO-lisenssiä.
- Julkaisuun tallennettavien metatietojen koneluettavuus on ehkä hankalin vaatimuksista, sillä ei ole aivan selvää, mitä sillä tarkoitetaan. Julkaisuutallennetaan julkaisuarkistoihin yleensä PDF-formaatissa, jonka osalta vaatimus tietojen koneluettavuudesta on erityisen ongelmallinen. Ei ole ylipäättään selvää miten tiedot kannattaisi tallentaa PDF-tiedostoon, ja lisäksi tiedostot tulevat usein valmiina esim. kustantajan palvelusta, eli niiden muokkaaminen voi olla vaikeaa. Yhtenä mahdollisena vaihtoehtona ehdotettiin vaadittujen tietojen lisäämistä PDF-tiedoston omiin metatietoihin. Ainakaan PDF-tiedostoja ei voi tallettaa kuvina, sillä ne eivät ole koneluettavia tai saavutettavuuslain vaatimusten mukaisia. Esiin nousi myös kysymys myös siitä, tukevatko e-kirjalukijoiden käyttämät formaatit kuten EPUB koneluettavuutta tai saavutettavuutta, mutta tästä ei ole varmuutta.
- Julkaisun pysyvä tunnistus eli PID (tieteellisillä julkaisuilla yleensä DOI tai URN) vaaditaan sekä alkuperäisestä julkaisusta että rinnakkaistallenteesta. Metadatasuosituksessa on kentät sekä alkuperäisen että rinnakkaistallenteen PID-tunnisteille. Tutkijoiden ORCID-tunnisteen ja organisaatiotunnisteen merkitseminen on suositus, ei pakollinen vaatimus. Organisaatiotunnisteen käyttö on vielä kehitteillä ja asiasta käydään keskustelua. Tiedeyhteisö suosii ORCID- ja ROR-tunnuksia ja kirjastomaailma ISNIä.
- Artikkelin julkaisuarkistoon tallennetun kopion versio (AAM eli Author Accepted Manuscript tai VoR eli Version of Record) on merkittävä metatietoihin. Metadatasuosituksessa on varattu kenttä versiotiedolle. Tässä yhteydessä keskusteltiin kansilehdestä, jolle vaaditut metadatatiedot voisi koota ja siten saada ne mukaan itse julkaisuun. Jyrki I. kertoi sen olevan yksi mahdollisuus, josta on maailmalla keskusteltu intohimoisesti jo pitkään, sekä puolesta ja vastaan. Kansilehtiä käytetään monissa julkaisuarkistoissa, mutta esim. Google Scholar ei suosittele niitä, sillä ne hankaloittavat indeksointia. Asiaa täytyy jossain tapauksessa selvittää.
- Julkaisun Open Access (OA) -statuksen esittämiseen riittää OpenAIREn suosituksen soveltaminen, ja metadatasuosituksessa on tähän tarkoitettu kenttä, jolle on määritelty OpenAIREn suosituksen mukaiset mahdolliset arvot.
- Julkaisun lisenssille on metadatasuosituksessa määritelty kenttä ja sille arvot.
- Tutkimuksen rahoittaja ja rahoitus herätti eniten keskustelua. Metadatasuosituksessa on kolme rahoittajaan viittaavaa kenttää, mutta ne eivät täysin kata kaikkia rahoittajien vaatimuksia, ja lisäksi esim. kenttä *dc.relation.projectid* on varattu ensisijaisesti EU-projekteille, eli siihen tallennettujen tietojen pitää olla tietyssä OpenAIREn määrittelemässä spesifissä formaatissa. Miika Nurminen kommentoi, että rahoittajatietojen merkintä metadataan on todellakin haaste, ja hän esitteli Jyväskylän yliopiston julkaisuarkistossa tähän käytetyt, tutkimustietojärjestelmästä siirtyvät kentät:

- *dc.relation.funder* (rahoittajan nimi)
- *dc.relation.grantnumber* (päättönumero, muut kuin EU-projektit)
- *dc.relation.projectid* # OpenAIREn EU-projektitunniste, esim. *info:eu-repo/grantAgreement/EC/H2020/688520/EU//TeSLA*
- *jyx.fundingprogram* # Rahoitusohjelma, käytetään rahoittajalogojen näyttämiseen
- *jyx.fundinginformation* # Acknowledgements-teksti tai muuten vapaamuotoisesti kuvattu tieto rahoituksesta

Keskusteltiin siitä, että rahoittajätietojen selvittäminen ei ole aina helppoa, ja prosesseissa on organisaatiokohtaisesti vaihtelua. Tietoja kerätään kysymällä tutkijoilta ja poimimalla niitä artikkeleista. Myös viitetietokantoja kuten Scopusta ja Web of Sciencea voi käyttää apuna ja tietoja haravoida niistä automaattisesti. Tietokantojen tiedot voivat kuitenkin olla ristiriitaisia, joten rahoitustiedot on syytä ainakin tarkistaa. Jyrki H. kertoi Akatemian olevan hyvin kiinnostunut rahoitustiedoista ja huomautti, että kansallisen tutkimuksen rahoitustietoja ei löydy kansainvälisistä viitetietokannoista. Akatemia on pitänyt asiaa esillä. Tarvetta standardisoimiseen tietojen merkinnässä selvästi on, ja yhteisiä menettelytapoja yritetään löytää.

Vahvat suositukset

Keskustelu siirtyi Plan S:n vahvoin suosituksiin, jotka eivät siis ole pakollisia. Todettiin, että monien suositusten osalta niiden käyttöönotto on realistista vasta hieman pidemmän aikavälillä. Suosituksia kohti on hyvä pyrkiä ja pitää ne mielessä, vaikka niiden toteuttamista ei vaaditakaan. Kansainvälisen ja kansallisen keskustelun seuraaminen ja siihen osallistuminen on suositeltavaa, nämä muuttuvat koko ajan, tähdensi Jyrki H.

Avoimesta rajapinnasta puhuttaessa vaihdettiin kokemuksia. Jyväskylässä on aiemmin ollut käytössä kustantajalta suoraan tuleva tiedostopalvelu (BioMed Central), mutta se oli ongelmallinen metadatan yhteensopimattomuuden takia. Yhteydet vaativat kustantajakohtaista räätälöintiä, mikä ei ole ongelmatonta. Käytännössä aineistojen siirto on tapahtunut SWORD-protokollan avulla.

Myös Helsingin yliopiston Heldassa on suoraan kustantajalta saatuja tiedostoja. Kustantajakohtaisten prosessien rakentaminen on työlästä, joten tilalle on kaivattu JISC:in Britanniassa rakentaman Publication Router-palvelun kaltaista keskitettyä palvelua, joka välittäisi eri kustantajien aineistoja mahdollisimman yhtenäisessä formaatissa.

Metadatan laadunvarmistuksessa ulkoisista lähteistä voisi ottaa Plan S:n listaamien ulkomaisten lähteiden rinnalle [kansallisen tutkimustietovarannon](https://www.tiedejatutkimus.fi/fi/) osoitteessa <https://www.tiedejatutkimus.fi/fi/>.

Rights retention strategy

Plan S:n sisältämän rights retention strategyn merkitys on se, että tekijänoikeus jää tekijälle tai hänen edustamalleen organisaatiolle, ja julkaisijan on hyväksyttävä se. Tämä on kirjattu rahoitusehtoihin ja niiden kautta tieto saadaan tutkijalle / kirjoittajalle. Tiedonvälityksessä ja ohjeistuksessa kirjastohenkilökunta on kuitenkin tärkeässä roolissa.

Julkaisun lisenssi

Lisensseihin kohdistuu vaatimuksia. Suositus on, että oletuslisenssi on CC-BY (Nimeä), ja lisenssistä käytetään uusinta versiota (esim. tällä hetkellä CC BY 4.0). Lisäksi CC BY-SA ja periaatteessa myös CC0 ovat mahdollisia. Tällä hetkellä on valmisteilla menettely, jolla tutkija voisi lisäksi pyytää poikkeuslupaa CC BY-ND -lisenssin käyttämiseen, kun siihen on hyvin perustellut syyt.

Pauliina Raento Suomen tiedekustantajien liitosta toi esiin, että avoimuuskaan ei ole ilmaista, vaan myös kustantajien toimintaedellytykset pitää pystyä turvaamaan. Kotimaisten kustantajien kannalta CC-lisenssien ongelmana on se, että ne eivät ole Suomessa käytetyn kollektiivilisenssimallin mukaisia, ja ne vaarantavat siten myös oikeudenhaltijoiden Kopioston kautta saamat korvaukset. Vaarana on kielen ja tieteenalojen eriarvoistuminen, minkä vuoksi keskusteluun olisi hyvä saada mukaan myös OKM.

Miten nopeaa avoimuutta vaaditaan?

Todettiin, etteivät rinnakkaistallenteet erilaisista käytännön syistä johtuen tule aina aivan välittömästi saataville edes siinä tapauksessa, ettei niillä ole kustantajan määrittelemää embargoaikaa. Jyrki H. kertoi, että välitöntä avoimuutta ei ole aikayksikkötasolla tarkasti määritelty, mutta tosiasia on, että nykyiset embargot ovat liian pitkiä. Akatemia ei aseta sanktioita, mutta voi pyytää tutkijaa hoitamaan avoimuuden kuntoon havaitessaan puutteita. Päämääränä on saada suomalainen tiedeyhteisö toimivaksi niin, että tutkijoiden ja myös muiden on helppo tehdä työtään ja tutkimuksen tulokset saadaan nopeasti yhteiskunnan käyttöön.

Seuraavat askeleet

Lopuksi todettiin, että keskustelu on ollut tarpeellista ja antoisaa, ja sen jatkamiselle on tarvetta. TSV:n Avoin tiede -koordinaatiotyö mainittiin yhtenä mahdollisena estradina jatkokeskustelulle, samoin Finn-ARMA-verkosto.

Suomen Akatemian järjestää 27.11. tilaisuuden, jossa Plan S:n vaatimuksia käsitellään yleisemmällä tasolla. Jyrki H. sanoi vievänsä esiin tulleet asiat eteenpäin Coalition S:lle ja palaavansa asiaan tilaisuuden järjestäjien kanssa. Seminaariin osallistujille lähetetään linkki palautekyselyyn, jonka täyttämistä toivotaan.