

Asiantuntijaryhmän kokous 5.12.2018

Aika: ke 5.12. klo 13-16

Paikka: nh. Juhani, OKM (Meritullinkatu 1)

Tallenne: <https://kannu.csc.fi/s/GxCR2eEo5koxTJQ>

Osallistajat:

- Opetusneuvos Jukka Haapamäki (ryhmän puheenjohtaja), opetus- ja kulttuuriministeriö
- Suunnittelija Janne Pölönen, Tieteellisten seuran valtuuskunta
- Tietojärjestelmäasiantuntija Jyrki Ilva, Kansalliskirjasto
- Tietoasiantuntija Anna-Kaarina Linna, Turun yliopisto
- Tietoasiantuntija Leena Huiku, Aalto-yliopisto
- ~~Julkaisu-suunnittelija Tarja Laakkonen, Laurea ammattikorkeakoulu ESTE~~
- Asiantuntija Timo Hellgren, Teknologian tutkimuskeskus VTT
- Tiedeasiantuntija Otto Auranen, Suomen Akatemia
- ~~Business Finland~~
- Projektipäällikkö Joonas Nikkanen, CSC - Tieteen tietotekniikan keskus
- Koordinaattori Minna Ala-Mantila, CSC - Tieteen tietotekniikan keskus

Asialista/muistio

1. Kokouksen avaaminen / ilmoitusasiat
 - a. Puheenjohtaja Jukka Haapamäki avasi kokouksen klo 13.09
 - b. Ei ilmoitusasioita
2. Attribuuttipohjaisen luokittelun hahmotelma
 - a. [Luonnos attribuuttipohjaisesta julkaisutyypiluokittelusta](#)
 - b. **Päätettiin, että kommenttien pohjalta tehdään uusi versio sekä matriisi kaikkien mahdollisten yhdistelmin havainnollistamiseen ja nykyisen julkaisuluokittelu tyyppien vastaavuuteen.**
3. [Avoimen saatavuuden määritelmä tietomallissa](#) (Jyrki Ilva)
 - a. Nykyisen mallin esittely: julkaisun verkko-osoite, kustantajan version OA-status (kolme vaihtoehtoa), rinnakkaistallenne, rinnakkaistallenne verkko-osoite (kompleksisuutta embargoajan päättymisen merkitsemisessä käyttämisessä)
 - b. Ehdotus, julkaisun versiot jaoteltuna: 1. Julkaisu kustantajan palvelussa 2. Rinnakkaistallenne 3. Pre Print -versio
 - i. (1) *Julkaisu kustantajan palvelussa* (version tämän hetkinen OA-status, julkaisukanavan OA-status, kustantajan version lisenssi + mahdollisia artikkelimaksu, url)
 1. Oleellista: kerätään tietoa onko *tällä hetkellä* avoin. Saattaa muuttua tulevaisuudessa (muuttua avoimeksi tai suljetuksi). Julkaisukanavan OA-status rikastaisi tätä tietoa. Joustavuutta, ei sidoksissa yhteen OA-määritelmään.
 - ii. (2) *Rinnakkaistallenne* (on/ei + mahdollisia versio, lisenssi, embargon päättymisaika, url) (+kv OA-vaatimuksia: avoimen saatavuuden alkamispäivä + version tekninen formaatti)
 - iii. (3) *Pre-print versio* (on/ei + mahdollisia lisenssi, url + mahdollisti ajankohta)
 1. Mietittävä halutaanko preprint tietoa kerätä. Ei tällä hetkellä merkitystä rahoitusmalleissa tai OA-monitoroinnissa, lisätyötä tiedonkeruuseen - mutta joillain aloilla merkittävä.
 2. Pre-printit sekoitetaan usein rinnakkaistallennettuihin, mutta eri asia. Pre-print tallennetaan julkaisuarkistoon jälkikäteen, eikä vertaisarvioitu.
 3. Pre-printin määppäytyminen attribuutteihin ongelmallista, sekä pre-printtien mahdolliset suuretkin eroavaisuudet varsinaisesta julkaisusta.
 4. Onko organisaatioilla intressejä saada pre-printtejä omiin järjestelmiin? Kansallisen ohjauksen kannalta ei merkitystä. Ei julkaisun muoto vaan tila/versio? Jollain tieteenaloilla käyttö suurta, koska nopeat prosessit, tutkijoille tärkeää. Tarpeesta ja käytöstä voisi kysyä fysiikan paneelilta. Haapamäki luokittelisi omana julkaisumuotona, selkeyden kannalta. **Pölönen/TSV vie fysiikan paneelille, kysely pre-printtien merkittävytydestä ja koetaanko pre-print erillisenä muotona.**
 - c. Vaihtoehtoinen malli (abstraktimpi): kustantajan palvelussa julkaistu versio, rinnakkaistallenne ja pre-print tietyn version ominaisuuksia. Rajaton määrä versioita yhdestä julkaisusta. Samantapainen kuin Puresa.
 - d. Miten paljon julkaisukanavaa koskevia tietoja kannattavaa tallentaa - onko riittävää että keskitetysti Jufon julkaisukanavalistassa, vai onko oltava myös julkaisukohtaisissa tiedoissa?
 - i. Mahdollisesti kerättävistä tiedoista kanavan oa-status tärkein.
 - ii. Myös rinnakkaistallennuspolitiikka, tosin tässä ei juurikaan hyviä vaihtoehtoja, Sherpa/Romeo värikoodit ei toimivia tai luotettavia. Yksi mahdollinen jatkoselvitys, tunnistaa lehdet/kanavat jotka sallivat vertaisarvioitujen käsikirjoitusversion välittömän rinnakkaistallennuksen. Tutkimusrahoittajien tulisi panostaa resursseja kansainvälisen palvelun tietopohjan parantamiseen.
 - iii. Vaihtoehtoisesti rinnakkaisen kansallisen luokituksen rakentaminen.
 - iv. Listahinnat muuttuvat usein, ei merkitystä. Todellinen maksettu hinta kiinnostavampi. Tietojen kerääminen ja ylläpito mietittävä.
 - v. Kaikki em. tiedot sellaisia, jotka voivat muuttua. Tiedot pitäisi sitoa vuoteen, vuosiversioituja tietoja. DOAJ status ja Sherpa/Romeo päivitetyt tiedot tulevat automaattisesti, jufolle.
 - vi. Miten tutkijat saavat tiedot julkaisukanavan tulevaisuuden OA-statuksen muutoksista? Myös varoituksia maksullisuuksista, tietoa tarpeellisten tunnisteiden löytymisestä jne. Nyt näitä listoja useampia, tietoja ei jaeta tehokkaasti. Jufo synkkaa kv-predator listojen kanssa. Todettiin, että julkaisufoorumin nykyinen 0-1 jaottelu riittävä, eikä uusia varoitustietoja tarvita.
 - vii. Vuosiversiointitietoa OA-statuksista ei lisätty julkaisufoorumin kantaan, vaikka tiedonkeruun kannalta olisi hyödyllistä. Mahdollisesti automatisointi DOAJ ja Sherpa/Romeo hälyttimillä.
 - viii. Haapamäki: OA-status tulossa rahoitusmalliin, joten sen kerääminen selkeämpää keskitetysti lähteenä tulevana tietona. Pölönen: tulisi tehdä tarkastelu, jossa katsotaan kuinka moni niistä julkaisuista jotka tiedonkeruuseen nyt merkitty avoimessa kanavassa julkaistuiksi on vielä DOAJ-indeksoiduissa julkaisuissa julkaistuja, sekä, onko paljon sellaisia artikkeleita, joita ei ole merkitty tiedonkeruussa avoimeksi vaikka julkaistu DOAJ-indeksoiduissa lehdissä. Myös tarkistaminen Sherpa/Romeo rinnakkaistallennetuiksi merkittyjen julkaisujen osalta onko lehti todellisuudessa rinnakkaistallentamisen mahdollistava ja toisinpäin. Rinnakkaistallentamisen tarkastamisessa yhteistyötä yliopistokirjastojen kanssa, tai tiedonkeruun kautta.

- ix. Haapamäki: OA-tietojen keräämien keskitettyyn julkaisukanavaan järkevää. Muut tukitietoja. Ilva: huomioitavaa, että kaikki julkaisukanavat eivät keskitetyssä mallissa mukana (esim. niin uusia), joten mietittävä keskitetyn mallin käytettävyyttä. Haapamäki: jos tieto keskitetyltä listalta niissä tapauksissa kun mahdollista.
 - x. Kannatusta myös julkaisukohtaisesti OA-tiedon lisäämiselle, ja siitä keskitetylle listalle heijastaminen. Eri OA-tasoja/kaksi eri OA-näkökulmaa: 1) onko OA ja 2) millä tavalla OA.
 - e. Olisi hyödyllistä saada yhteiseen käyttöön kirjastojen keräämää tietoa eri lehtien OA-statuksesta.
 - f. Yhteenvedettynä, ensimmäisen ja toisen mallin erona se, että ensimmäisessä mallissa oletetaan että erilaisia versioita olemassa, toisessa geneerisempi lähestyminen jossa erilaisia versioita ja erilaisia ominaisuuksia. Enemmänkin tekninen ero.
 - g. Keskeisenä kysymyksenä, voiko keskitettyyn kanavaan luottaa julkaisukanavan OA-statusen tiedon saamisessa, vai tarvitaanko tieto julkaisukohtaisesti.
 - i. Keskitetty kanava apuvälineenä, josta voidaan tarkistaa ja jota voidaan tarvittaessa korjata, vai luotetaanko sellaisenaan?
 - ii. Puresse ei tietoa onko hybridijulkaisu, ainoastaan onko OA. Tieto avoimuuden muodosta pitäisi tulla jostain muualta. Luokituksen lisääminen tälle hankala, olisi helpompaa jos voisi tarkastaa jostain. Jos kanavan status tulisi muualta, yksinkertaisiksi tiedonkeruuta.
 - iii. Tiedonkeruuseen on OA/ei OA, ja jos OA, tarkempi tieto muualta. Mutta tämä vaatisi taas muutoksen tiedonkeruun ohjeisiin, joskin muutos yksinkertaisempaan suuntaan... Huomioitava että kaikille tätä tietoa ei kuitenkaan saada, joten jokatapauksessa, tieto on löydettävä lähtöorganisaatiosta.
 - iv. Nykyisellään menetetään tieto esimerkiksi viivästetystä OA:sta, kun vain senhetkinen OA-status.
 - v. Jos tietoja saadaan keskitettyyn palveluun, voidaan tarkistaa a) nykyistä tietoa ja b) luoda tiedoista viivästetty OA-kategoria.
 - vi. Ajantasaisen tiedon ylläpitäminen olisi haasteellista, koska tiedot muuttuvat niin usein. Olisi ainakin oltava rajapinta, josta tietoa (keskitettyä listaa) mahdollista päivittää.
 - vii. Haapamäen ehdotus: **pysytään toistaiseksi OA-statusen osalta nykyisessä, ilmoitetaan artikkelitasolla. Kerätään tietoja keskitetysti julkaisukanavatietokantaan, josta organisaatioiden käyttöön, mahdollistaisi mm. viivästetyn avoimen tiedon ja nykyisen tiedon tarkistuksen.**
 - viii. Ilva: tulevaisuudessa DOAJ ja Unpaywall, jotka toimivat toisenlaisella logiikalla. Jos halutaan jossain vaiheessa purkaa osiin Suomessa nykyisin käytössä oleva järjestelmä, on otettava kantaa siihen mitä tehdään viivästetyn OA:n kanssa, oltava eroteltuna.
 - h. Aikataulusta: ei uuteen tiedonkeruuseen OA-statusen osalta, mutta julkaisutietopalvelun osalta aletaan rakentamaan rekisteriä, ja julkaisutietopalvelussa olevia tietoja voidaan alkaa rikastamaan. rekisterin keräyksen aloitus hetimiten, mutta tietomallia ei vielä.
 - i. TUHA-tiedonkeruuryhmässä pyydetty täydennyksenä käsikirjaan selventävää merkintää rinnakkaistallenteesta, ehdotettu että arxivin tallennetun jo julkaistun julkaisun voisi ilmoittaa rinnakkaistallenteena. Voidaanko yksittäiselle tieteenalalle antaa poikkeuslupaa?
 - i. Ilva: periaatteessa ei mitään epäselvää, Arxiv lähtökohtaisesti rajapinta-arkisto, eli suurin osa julkaisuista pre-printtejä (eli ei vertaisarvioitu), jos julkaisu vertaisarvioidaan, tutkijan tulisi tallentaa erikseen vertaisarvioitu versio. Vertaisarvioidun version voi ilmoittaa rinnakkaistallenteena myös tiedonkeruussa.
 - ii. Haapamäki: **selventävän merkinnän rinnakkaistallenteesta voi kirjoittaa auki tiedonkeruun käsikirjaan.**
 - j. Lisenssitieto: vaihtoehtoisiksi kirjava valikoima erilaisia lisensejä. Lisenssitietoa tullaan edellyttämään [Plan S](#):ssä. Tarjottavia vaihtoehtoja voi pohtia. Tätä tietoa ei keskitetystä rekisteristä.
 - i. Haapamäki: **lisenssitiedon lisääminen tietomalliin vapaaehtoisena kenttänä. Pre-print tulee mahdollisesti omana muotonaan.**
 - ii. Tietomallin kannalta mietittävä, onko julkaisutason asia vai voiko eri versioita ilmoittaa erikseen (joskin saattaa olla epätodennäköistä että monia tapauksia joissa useita eri lisensejä).
 - k. Artikkelimaksu: ei vielä päätöstä kansalliskirjastolta.
 - l. URL, tieto kerätään myös nykyään, ok.
 - m. Haapamäki: **OA-statusen osalta ei vielä tehdä muutosta tiedonkeruun ohjeistukseen.**
 - n. Rinnakkaistallenteen...
 - i. ... versio tällä hetkellä Puresse pakollinen tieto. Haapamäki: **luodaan koodisto rinnakkaistallenteelle, pohjalle Puren malli.**
 - ii. **...lisenssi otetaan tietomalliin**
 - iii. **...embargon päättymisaika otetaan tietomalliin**
 - iv. **...url otetaan tietomalliin**
 - v. ...arkiston status (kuuluu Plan S linjauksiin)? Lähes kaikki hyväksytään, kunhan päivitetty tieto.
 - vi. ...sosiaalisen median palvelut (mm. ResearchGate jne), poissuljettu OA-määritelmästä, tämän voisi merkitä myös ohjeisiin.
 - vii. ...avoimen saatavuuden alkamispäivä - onko sama kuin embargon päättymispäivä / alkamispäivä jos ei embargoa. **Ei lisätä.**
 - viii. ...versio tekninen formaatti - ei vielä ajankohtainen, mutta Plan S:ssä. **Ei lisätä.**
 - o. **Pre-print tarpeen käsittelystä jatketaan myöhemmin.** Haapamäen ehdotus: pre-print muotona.
4. Meriittiluokitus (*ei käsitelty*)
 5. Pilotointi ja aikataulu
 - a. Meriittiluokituksen pilotointi (*ei käsitelty*)
 - b. Attribuuttipohjaisen tietomallin pilointi
 - i. Ensimmäisenä mahdollisena vaihtoehtona, että vuoden 2020 tietoja tuotetaan attribuuttipohjaisena. Pure-käyttäjille arvioidaan mahdolliseksi aikatauluksi, Converis haastavampi. Ehtona että CSC pystyy myös purkamaan, nykyisestä julkaisutyypiluokituksesta attribuutteihin. Pidetään siirtymäaika kokonaan attribuutteihin siirtymiselle. **Tavoitteena että 2020 tiedot mahdollista tuoda attribuutteina, tammikuun kokoukseen aikataulu mikä arvio olisi mahdollista.**
 - c. Muut asiat
 - i. **Sovitaan tavoitteeksi että organisaatioille lähetään matriisiluonnos viikkoa ennen seuraavaa kokousta.**
 - d. Kokouksen päättäminen
 - i. Puheenjohtaja päätti kokouksen klo 16.00