

Hybridiartikkelien tunnistaminen edellyttää yhteisiä metatietokäytäntöjä

Fakta Julkaistu 16.01.2018 klo 10:12 Kirjoittaja Juuso Ala-Kyyny ja Kimmo Koskinen

Hybridiartikkelien julkaisemisen yleistymisen viime vuosina on lisännyt tarvetta hybridiartikkelien tunnistamiseen ja kartoittamiseen. Kokonaiskuva hybridien määrästä olisi tärkeää erityisesti yliopistojen ja tutkimusrahoittajien näkökulmasta, koska hybridijulkaisemisen osuus avoimesta julkaisemisesta ja siihen liittyvistä kustannuksista on merkittävä. Myös niin sanotun “double dipping” -ilmiön taustalla on hybridien tunnistamisongelma. Toistaiseksi ratkaisut hybridiartikkelien tunnistamiseen puuttuvat. Tässä artikkelissa käymme läpi hybrideihin liittyvää ongelmakenttää, ja esitämme vaihtoehtoja hybridien tunnistamiseksi.

Hybridiartikkelilla tarkoitetaan tilausmaksullisessa (subscription) lehdessä julkaistua avoimesti saatavilla olevaa (open access) artikkelia. Artikkelit on usein lisensoitu Creative Commons -lisenssillä. Kustantaja perii hybridiartikkelin julkaisemisesta eli artikkelin avaamisesta kirjoittajamaksun (APC, article processing charge) samaan tapaan kuin avoimissa lehdissä (gold OA). Hybridiartikkelien kirjoittajamaksu on OpenAPC-palvelun keräämien tietojen perusteella keskimäärin noin 2500 euroa.

Hybridilehtien määrä on kasvanut 2010-luvulla. Björkin (2017) tutkimuksen mukaan suurimpien tiedekustantajien tarjoamien hybridilehtien määrä on viisinkertaistunut 2000:sta 10000:een vuosina 2009–2016. Samana aikana julkaistujen hybridiartikkelien määräkin on noussut 8000 artikkelista 45000 artikkeliin. Luvut ovat arvioita, koska hybridiartikkelien kartoittaminen on vaikeaa – tämä vaikeus on tämänkin artikkelin aiheena. Avoimen julkaisemisen hybridimuoto on ollut erityisesti suurten kustantajien mieleen, ja valtaosa tärkeimmistä tiedelehdistä tarjoaa hybridivaihtoehdon. Kustantajien kannalta hybridimalli on houkutteleva taloudellisessa mielessä. Tutkimusrahoituksessa avoimen julkaisemisen kustannukset on otettu entistä paremmin huomioon, mikä sekin on Björkin mukaan edistänyt hybridijulkaisemista.



Kuva: Robert Cudmore / Flickr CC-BY-SA 2.0

Hybridien tunnistamisen merkitys

Hybridiartikkelien tunnistamiselle on useita syitä. Tunnistaminen mahdollistaa hybridiartikkelien tallentamisen julkaisuarkistoon ja auttaa hallitsemaan nykyistä paremmin avoimeen julkaisemiseen liittyviä kirjoittajamaksuja. Hybridiartikkelien tunnistaminen myös täydentää tutkimustietojen vuosiraportointia. Käyttäjän kannalta hybridien tunnistamisen merkitys liittyy tiedon saatavuuteen.

Hybridiartikkelit ovat avoimia artikkeleita, joten hybridiartikkeleista voidaan tallentaa julkaisuarkistoon kustantajan versio. Hybridiartikkelien tallentaminen julkaisuarkistoon on tärkeää, vaikka ne ovatkin avoimesti saatavilla. Ennen kaikkea se takaa tutkimustiedon pitkäaikaissaatavuuden. Kaupallisia kustantajia tulee ja menee, ja omistusmuutoksissa hybridiartikkelien asema ei ole selvä. Kun artikkeli on asianmukaisesti tallennettu julkaisuarkistoon, tutkimustiedon saatavuus on vakaammalla pohjalla ja tallentaminen tukee organisaation muistia.

Hybridiartikkelien kartoittaminen on tärkeää myös yliopiston tutkimustietojen raportoimisen kannalta. Opetus- ja kulttuuriministeriö kysyy hybriditietoja vuosiraportoinnin yhteydessä, ja on mahdollista, että jossain vaiheessa nämä tiedot vaikuttavat myös yliopistojen rahoitukseen.

Yliopiston näkökulmasta tieto hybridiartikkeleista liittyy keskeisesti myös kirjoittajamaksujen hallintaan. APC-maksujen hallinnointi edellyttää tietoja kaikista yliopiston maksamista OA-maksuista. Hybridiartikkelien kartoittaminen liittyy myös niin sanotun kaksoismaksatuksen (double dipping) välttämiseen. Hybridimallia on kritisoitu erityisesti siitä, ettei avattuja hybridiartikkeleita aina huomioida organisaatioilta perittävissä tilausmaksuissa. Tällöin yliopistot saattavat maksaa kahteen kertaan – sekä kirjoittaja- että tilausmaksuna – samasta aineistosta.

Hybridien tunnistaminen on tärkeää tietysti myös käyttäjien kannalta, vaikka hybridistatus sinänsä on tiedonhakijalle toissijainen seikka – olennaista on artikkelin avoimuus. Tunnistamisongelmien takia avoin hybridiartikkeli jää kuitenkin usein käytännössä maksumuurin taakse. Avoimeen julkaisemiseen perinteisesti liittyvää ajatusta tiedon maksuttomasta, vapaasta saatavuudesta verkon välityksellä olisikin hyvä ehkä laajentaa: tiedon tulisi olla myös riittävän *helposti* käyttäjän saatavilla, jotta avoimuus todella toteutuu. Avointen julkaisujen löytämiseen liittyvä näkökulma tuo OA-artikkeleiden tunnistamiseen hakupalvelu-ulottuvuuden. Hybridiartikkelit tulisi voida voida tunnistaa yksinkertaisesti metatietojen perusteella, jotta tiedonhakujärjestelmät osaisivat tarjota niitä käyttäjille tehokkaalla tavalla (Bullock 2017). Hybridiartikkelien sisällyttäminen kirjastojen tarjoamiin tutkimuksen tietopalveluihin on haasteellista, mutta tähän suuntaan ollaan pyrkimässä (Pennington 2016).

Hybridien tunnistamisen vaikeudet ja tunnistamisen perusmallit

Hybridiartikkelien tunnistaminen ja kartoittaminen olisi edellä kuvattujen syiden vuoksi tärkeää, mutta nykytilanteessa se on vaikeaa. Hybridiartikkelien tunnistamiseen ei ole olemassa mitään vakiintunutta käytäntöä, ja jokaisella kustantajalla on oma tapansa ilmaista hybridit. Yleensä hybridiartikkeleita ei ole mahdollista hakea kustantajan hakupalvelusta suoraan, vaan artikkelin hybridistatus on pääteltävä artikkelin avoimuuden ja lehden statuksen (tilausmaksullinen vai avoin lehti) perusteella. Tämäkään ei vielä riitä. Se, että tilausmaksullisessa lehdessä on avoin artikkeli, ei automaattisesti tarkoita hybridiartikkelia, jos lisenssitieto puuttuu. Kyseessä voi olla hybridin sijaan niin sanottu bronze OA -artikkeli, jonka kustantaja on omasta aloitteestaan avannut joko pysyvästi tai väliaikaisesti. Artikkelin lisenssitieto (Creative Commons tai vastaava) on ratkaiseva, koska se vahvistaa artikkelin avoimuuden. Lisenssimerkintä voi olla artikkelin verkkosivulla tai pdf-tiedostossa.

Hybridiartikkelien tunnistaminen voidaan jakaa manuaaliseen ja automaattiseen tunnistamiseen. Edellä on kuvattu manuaalisen tunnistamisen prosessia. Helsingin yliopiston kirjastossa tehtiin kesän ja syksyn aikana hybridien manuaalista tunnistustyötä, ja keskeisiä vaiheita tässä Helsingin yliopiston tutkijoiden julkaisemien hybridiartikkelien manuaalisessa tunnistamisessa olivat (1) affiliaation määrittäminen, (2) avointen artikkelien löytäminen ja (3) avointen artikkelien tunnistaminen hybrideiksi toisaalta lehden statuksen, toisaalta artikkelin lisenssitietojen perusteella. Apuvälineitä tunnistamiseen on olemassa, esimerkiksi OA-lehtiä on koottu DOAJ-tietokantaan. Myös kustantajien sivuilta voi löytyä lista OA-lehdistä tai tilausmaksullisista lehdistä. Tunnistustyö ei silti ole yksinkertaista. Hakukäytäntöjen lisäksi myös käsitteet saattavat olla hämääviä, sillä ”free access” tai ”full access” eivät useinkaan tarkoita artikkelin avoimuutta samassa merkityksessä kuin ”open access”. Hybridiartikkelien manuaalinen tunnistaminen onkin aikaavievää käsitelyä, jossa hakustrategiat vaihtelevat huomattavasti kustantajasta toiseen.

Tieteellisen kirjallisuuden keskeisistä viiteindekseistä Web of Science (WoS) hakumahdollisuudet täydentyivät merkittävästi loppuvuodesta 2017, kun artikkeleita pystyi suodattamaan haussa myös avoimuustiedon perusteella. OA-vaihtoehtoina Clarivate Analyticsin WoS tarjoaa varsinaisten avointen artikkelien (”Gold”) ohella rinnakkaistallennettuja (”Green Accepted” ja ”Green Published”) artikkeleita. Hybridiartikkeleita haussa ei voi eritellä – ne löytyvät ”Gold”-artikkeleiden joukosta, ja tunnistaminen edellyttää tietoa lehden statuksesta ja artikkelin lisenssitiedosta. Kuvaavaa on WoS:n käyttämät nimitykset ”Gold Hybrid” ja ”Hybrid Gold OA”, joiden voi tulkita ilmaisevan hybridien erottelukseen liittyvää hankaluutta.

Tehokkaampi tapa hybridiartikkelien kartoittamiseen olisi niiden automaattinen tunnistaminen. Automaattisessa tunnistamisessa keskeistä olisi hybriditiedon löytyminen artikkelin metatiedoista. Hybridien merkitsemiseen ei kuitenkaan ole olemassa metatietostandardia, vaikka toimiva ratkaisu tähän on esitetty (Chumbe ym. 2015). Toinen esitetty ratkaisumalli OA-statusen metatiedoille on NISO-standardointiorganisaation Access and License Indicators -käytäntö. Kumpakaan näistä ratkaisuista ei toistaiseksi sovelleta laajasti.

Edellämainitun Web of Science -tietokannan OA-artikkeleiden tunnistus perustuu Impactstoryn kehittämään oADOI-palveluun (<https://oadoi.org/>), joka yrittää tunnistaa artikkelin OA-statusen käyttäen taustalla useita eri tietolähteitä, esim. CrossRef-palvelun tietoja. Open Access -tilan lisäksi oADOI pyrkii löytämään myös julkaisuarkistoihin tallennettuja versioita julkaisusta.

Toinen palvelu, joka seuraa hybridi-OA:n tilannetta on Göttingenin yliopistossa kehitetty Hybrid OA Journal Monitor (<https://najkoja.shinyapps.io/hybridoa/>). Se yhdistää tietoja CrossRef-palvelusta ja OpenAPC-palvelusta ja tarjoaa myös tietoja Creative Commons -lisenssien käytöstä. Tämäkään palvelu ei kuitenkaan auta tunnistamaan hybridejä artikkelitasolla. Toistaiseksi hybridiartikkelien automaattinen tunnistus on siis haaste, joka on vielä ratkaisematta.

Lisätietoja

- Björk B. (2017) [Growth of hybrid open access](#), 2009–2016. PeerJ 5:e3878
- Bullock, C. (2017) [Delivering Open](#), Serials Review, 43:3-4, 268-270
- Chumbe, S., Kelly, B. & MacLeod, R. (2015) [Hybrid Journals: Ensuring Systematic and Standard Discoverability of the Latest Open Access Articles](#), The Serials Librarian, 68:1-4, 143-155,
- Chumbe, S. S., MacLeod, R. A., & Kelly, B. (2015). [We should not light an Open Access lamp and then hide it under a bushel!](#) In B. Schmidt, & M. Dobrev (Eds.), *New Avenues for Electronic Publishing in the Age of Infinite Collections and Citizen Science: Scale*,

Openness and Trust: Proceedings of the 19th International Conference on Electronic Publishing (pp. 102-112). IOS Press. DOI: 10.3233/978-1-61499-562-3-102

- [Hybrid OA Journal Monitor](#)
- [OpenAPC](#)
- [NISO RP-22-2015 Access and License Indicators](#)
- Pennington, B. (2016) [Where Is That? The Challenges of Including Hybrid Journal Articles in the Library Research Process](#), *Serials Review*, 42:4, 306-310,