

## OKM:n tiedonkeruun avoin saatavuus -tieto

### 1. Kerättävä tieto ja tehtävänanto

OKM:n tiedonkeruun avoin saatavuus -tieto koodataan tällä hetkellä seuraavan ohjeistuksen mukaisesti (Yliopistojen tiedonkeruun käsikirja 2014, s. 33-34):

#### **0 ei Open access -julkaisu**

#### **1 Open access -julkaisu**

- julkaisu on ilmestynyt lehdessä tai sarjassa, jonka kaikki artikkelit tai osajulkaisut ovat pysyvästi ladattavissa ja luettavissa Internetin kautta ilman korvausta.

#### **2 Open access -julkaisu, rinnakkaistallennettu**

- tilausmaksullisessa lehdessä tai muussa lukijoiden kannalta maksullisessa teoksessa ilmestyneen julkaisun pysyvästi saatavilla oleva open access -kopio, joka on rinnakkaistallennettu esim. organisaation omaan tai tieteenalakohtaiseen digitaaliseen arkistopalveluun. Julkaisu voi olla joko ns. kustantajan versio tai tutkijan oma viimeinen (vertaisarvioitu) versio ja se voi olla saatavissa välittömästi tai kustantajan määrittämän kohtuullisen mittaisen (esim. puoli vuotta) embargoajan jälkeen. Julkaisu voi olla myös tiedeyhteisön käyttöön ennen virallista julkaisua annettu pre print -versio, jos se on pysyvästi saatavilla organisaation tai tieteenalan digitaalisessa arkistopalvelussa. Otetaan huomioon myös ne yksittäiset julkaisut, jotka ovat avoimesti saatavilla, vaikka itse sarja ei olisikaan avoimesti saatavilla.

#### **9 ei tietoa onko julkaisu Open access -julkaisu.**

Vastausvaihtoehtoja on siis neljä, ml. ”ei tietoa”. Open access -julkaisut on jaettu kahteen ryhmään, joista ensimmäiseen (”1 Open access -julkaisu”) kuuluvat open access -lehdissä ja sarjoissa ilmestyneet julkaisut, jälkimmäiseen (”2 Open access -julkaisu, rinnakkaistallennettu”) taas sekä julkaisuarkistoon rinnakkaistallennetut julkaisut että satunnaiset muut avoimesti saatavilla olevat julkaisut.

Seuraavassa käydään lyhyesti esimerkkien avulla läpi sitä, miten laadukkaita korkeakouluilta kerätyt koodiston mukaiset tiedot ovat olleet ja millaisia muita haasteita open access -julkaisuja koskevien tietojen keräämiseen liittyy. Selvitys keskittyy pääosin yliopistojen tietoihin. Lopuksi esitetään joitakin toimenpide-ehdotuksia tiedonkeruun kehittämiseksi.

Open access -tietoa ei sen oletetun heikon laadun vuoksi ole toistaiseksi käytetty Juuli-portaalissa, joten analyysi on tehty CSC:n wikissä julkaistun vuosien 2011-2013 raakadatan<sup>1</sup> ja Vipunen-portaalin sisältämien tietojen pohjalta.

### 2. Vuosien 2011-2013 julkaisutiedot

Yliopistojen vuosien 2011-2013 julkaisutiedoissa avoin saatavuus -tiedon eri vastausvaihtoehdot jakautuvat Vipusen mukaan seuraavasti:

<sup>1</sup> <https://confluence.csc.fi/display/suorat/Julkaisutiedonkeruu+-+korkeakoulujen+raportoimat+julkaisut>.

	määrä	osuus
Ei open access -julkaisu	41905	38,4 %
Open access -julkaisu	14951	13,7 %
Open access -julkaisu, rinnakkaistallennettu	2854	2,7 %
Ei tietoa	49480	45,3 %

Eri tietojen osuudet ovat vaihdelleet jonkin verran ajanjakson aikana, mutta muutokset eivät ole olleet kovin suuria. ”Ei tietoa”-vastausten määrä on kolmen vuoden aikana pienentynyt hieman, mutta niiden osuus on edelleen huomattava.

## 2.1. Tietolähteet

”Ei tietoa” on selkeästi yleisin avointa saatavuutta koskeva vastaus, mitä voi jo sinällään pitää osoituksena tietoihin liittyvistä ongelmista. Tämän kategorian ja muiden kategorioiden osuudet raportoiduissa tiedoissa vaihtelevat merkittävästi yliopistoittain. ”Ei tietoa”-vastausten osuus on erityisen suuri Helsingin yliopistossa ja Aalto-yliopistossa. Se on julkaisumäärään nähden huomattava myös mm. Turun yliopistossa, Åbo Akademiassa ja Vaasan yliopistossa. Helsingin yliopiston osalta tämä selittyy sillä, että kategoriaa ”Ei open access -julkaisu” ei ole käytetty lainkaan, eli siihen mahdollisesti kuuluvat julkaisut ovat kaikki ”ei tietoa”-kategoriassa. Toisaalta, muista yliopistoista Aalto-yliopisto ei puolestaan ole käyttänyt lainkaan ”Open access -julkaisu, rinnakkaistallennettu”-kategoriaa. Tällaiset säännönmukaiset eroavaisuudet selittyvät suurelta osin yliopistojen sisäisten tiedonkeruuprosessien eroilla.

### **Esimerkki: Avoin saatavuus -tiedon tuottaminen Helsingin yliopistossa<sup>2</sup>**

Helsingin yliopiston TUHAT-järjestelmän syöttölomakkeessa ei ole omaa kenttää open access -tiedolle. Julkaisutietoja tallentavat tutkijat eivät siis voi itse liittää julkaisunsa yhteyteen tietoa siitä, onko kyseessä open access -julkaisu vai ei. Julkaisu on tosin mahdollista rinnakkaistallentaa TUHAT:in kautta yliopiston Helda-julkaisuarkistoon.

Tutkijan syöttämien tietojen sijaan open access -julkaisuja tunnustetaan Helsingin yliopistossa keskitetysti olemassa olevien tietovarantojen pohjalta. Tunnistamista tehdään toisaalta sen perusteella, ovatko julkaisukanavat mukana Directory of Open Access Journalsin (DOAJ) lehtilistalla, toisaalta sen perusteella onko julkaisut rinnakkaistallennettu Heldaan. Jos julkaisua ei näiden kahden lähteen perusteella voida tunnistaa OA-julkaisuksi, sen statukseksi merkitään ”ei tietoa”.

Helsingin yliopiston käyttämä kahteen tietolähteeseen perustuva prosessi tunnistaa melko suuren osuuden tieteellisissä open access -lehdissä ilmestyneistä artikkeleista ja yliopiston omaan julkaisuarkistoon tallennetut artikkelit. DOAJ:hin listaamattomissa julkaisukanavissa (ml. hybridijulkaisukanavat, kirjasarjat ja konferenssijulkaisut, ei-tieteelliset julkaisukanavat) ilmestyneitä tai muihin kuin yliopiston omaan julkaisuarkistoon tallennettuja artikkeleita ei tässä prosessissa sen sijaan tunnisteta.

DOAJ:n (<http://doaj.org/>) kaltaiset ulkoiset tietolähteet ovat hyödyllisiä apuvälineitä tiedonkeruussa, sillä niiden perusteella on mahdollista tunnistaa automaattisesti suhteellisen suuri joukko open access -julkaisuja. Toisaalta ne kattavat jo lähtökohtaisesti vain osan tapauksista - esim. DOAJ:ssa on mukana ainoastaan tietyt suhteellisen tiukat tekniset ja laadulliset kriteerit kattavia tieteellisiä

<sup>2</sup> Prosessin kuvaus perustuu Aija Kaiteralta saatuihin tietoihin.

julkaisukanavia, eikä se ole niidenkään osalta kattava. Mm. suuri osa suomalaisista open access -lehdistä puuttuu toistaiseksi sen listalta.

DOAJ:n ulkopuolelle jääneissä julkaisukanavissa ilmestyneiden open access -julkaisujen tunnistaminen on huomattavasti hankalampaa. Hybridijulkaisukanavissa ilmestyneiden julkaisujen kohdalla open access -tieto on toistaiseksi pakko lisätä julkaisu kerrallaan, vaikka tähän saattaa olla tulevaisuudessa luvassa parempia ratkaisuja (kustantajien yhteistyössä rakentaman CHORUS-palvelun - <http://www.chorusaccess.org/> - voi olettaa jossain vaiheessa tarjoavan tällaisia tietoja kootusti). Ei-tieteellisten julkaisukanavien osalta valmiit tietolähteet lienevät jatkossakin puutteellisia, eli niiden kohdalla open access -tieto pitää todennäköisesti lisätä julkaisun yhteyteen sen tietojen tallennuksen yhteydessä.

Myös rinnakkaistallennettujen julkaisujen tunnistaminen on toistaiseksi haastavaa. Toisin kuin esim. Ruotsissa, Norjassa ja Tanskassa, Suomessa yliopistojen julkaisuarkistoja ei ole tällä hetkellä kytketty tutkimustietojärjestelmistä tehtyyn keskitettyyn tiedonkeruuseen, eli niihin tallennettuja julkaisuja koskevien tietojen tuominen julkaisudataan on kunkin raportoivan organisaation vastuulla. Tietoja olisi kuitenkin mahdollista haravoida myös keskitetysti ainakin kotimaisista julkaisuarkistoista, mahdollisesti myös keskeisistä kansainvälisistä julkaisuarkistoista.

## 2.2. Open access -julkaisut

Vipusen mukaan yliopistojen raportoimissa vuosien 2011-2013 tiedoissa on yhteensä 14951 julkaisua, joiden on ilmoitettu ilmestyneen open access -julkaisukanavissa. Tätä ei kuitenkaan ole helppoa verifioida, sillä lähes kolmasosalle näistä julkaisuista ei ole ilmoitettu lainkaan verkko-osoitetta, ja osassa muistakin tapauksista osoitteissa on selkeitä virheitä. Linkit saattavat myös viedä itse julkaisun sijasta verkkolehden etusivulle, jonkin yliopiston laitoksen sivuille tai joissakin tapauksissa myös julkaisuarkistoihin. Mukana on myös erilaisia eksoottisia tapauksia, ml. linkkejä esim. Googlen hakutuloksiin tai Dropbox-palveluun.

Yhteensä 2331:lle tähän kategoriaan kuuluvalle julkaisulle oli ilmoitettu dx.doi.org-palvelun kautta resolvoituva DOI-pohjainen osoite, minkä lisäksi datassa oli jonkin verran myös kustantajien palveluiden kautta resolvoituvia DOI-osoitteita. Käytännössä DOI:n perusteella tavoitettavien julkaisujen osuus on kuitenkin suurempi, sillä osalle muistakin julkaisuista on raportoitu DOI-tunniste erikseen sille varatussa kentässä, vaikkei siitä ole muodostettu DOI-pohjaista osoitetta.

Muista pysyvistä osoitteista URN-pohjaisia osoitteita on ilmoitettu 305 julkaisulle, eli niiden osuus on myös melko suuri. Etenkin ei-tieteellisille julkaisuille on kuitenkin ilmoitettu pääosin muita kuin pysyviä osoitteita. Yliopistojen lisäksi tämä koskee myös ammattikorkeakoulujen dataa, jossa suurin osa raportoiduista open access -julkaisuista on muita kuin varsinaisia tieteellisiä julkaisuja.

Tietojen yleistä luotettavuutta on tässä yhteydessä mahdollista selvittää ainoastaan otosten pohjalta. Seuraavassa yliopistojen vuosien 2011-2013 julkaisudatasta on poimittu neljän erilaisilla periaatteilla toimivan julkaisukanavan (tai tietyn kustantajan julkaisemien lehtien) tiedot. Kuten tiedoista käy ilmi, julkaisukanavakohtaiset erot tietojen luotettavuudessa ovat huomattavan suuria.

**Public Library of Science (PLOS)** on yksi tunnetuimmista open access -kustantajista. Se julkaisee sekä mailman suurimmaksi tieteelliseksi julkaisukanavaksi kasvanutta PLoS One -megalehteä että muutamaa muuta perinteisempää vertaisarviointiprosessia hyödyntävää open access -lehteä.

Yliopistoilta kerätyissä julkaisutiedoissa PLoS:in lehdissä ilmestyneiden 1107 artikkelin OA-status on määritelty seuraavasti:

	määrä	osuus
Ei open access -julkaisu	24	2,2 %
Open access -julkaisu	913	82,5 %
Open access -julkaisu, rinnakkaistallennettu	5	0,5 %
Ei tietoa	165	14,9 %

Kaikkien PLoS:in lehdissä ilmestyneiden artikkelien pitäisi olla siis open access -julkaisuja. Käytännössä lehtien status on jäänyt suurelta osin tunnistamatta Turun yliopistossa, Åbo Akademiassa ja Aalto-yliopistossa. Väärät tiedot ("Ei open access -julkaisu") ovat peräisin Itä-Suomen yliopistosta, Tampereen teknillisestä yliopistosta ja Oulun yliopistosta.

American Physical Society (APS) julkaisemat **Physical Review** -lehdet kuuluvat fysiikan alan merkittävimpiin kansainvälisiin julkaisukanaviin. Niiden osalta julkaisuille raportoidut OA-statukset näyttävät tältä:

	määrä	osuus
Ei open access -julkaisu	311	27,9 %
Open access -julkaisu	18	1,6 %
Open access -julkaisu, rinnakkaistallennettu	121	10,9 %
Ei tietoa	665	59,6 %

APS:n lehdet eivät ole open access -julkaisukanavia, eli niiden osalta oikeita vastauksia olisivat joko "ei open access -julkaisu" tai "open access -julkaisu, rinnakkaistallennettu". Ylivoimaisesti suurin kategoria on kuitenkin "ei tietoa", mikä selittyy suurelta osin Helsingin yliopiston ja Aalto-yliopiston raportoimilla tiedoilla. Rinnakkaistallennetuiksi merkityt artikkelit ovat pääosin Jyväskylän yliopistosta. Rinnakkaistallenteiden osoitteita ei ole yleensä raportoitu, mutta koska kyse on fysiikasta, saattaa olla, että ne on tallennettu tieteenalan omaan ArXiv-julkaisuarkistoon.

**Historiallinen aikakauskirja** on yksi keskeisistä kotimaisista humanistisen alan julkaisukanavista. Sen osalta julkaisutiedoissa raportoidut avoin saatavuus -tiedot näyttävät tältä:

	määrä	osuus
Ei open access -julkaisu	127	61,4 %
Open access -julkaisu	1	0,5 %
Open access -julkaisu, rinnakkaistallennettu	2	1,0 %
Ei tietoa	77	37,2 %

Historiallinen aikakauskirja ei ole open access -julkaisu, eli sen artikkelien tulisi sijoittua kategorioihin "ei open access -julkaisu" tai "open access -julkaisu, rinnakkaistallennettu". Yli kolmasosa artikkeleista on kuitenkin jäänyt "ei tietoa"-kategoriaan, mikä johtuu suurelta osin Helsingin yliopiston raportoimista tiedoista.

Terveyden ja hyvinvoinnin laitoksen (THL) suojissa toimiva **Yhteiskuntapolitiikka**-lehti ilmestyy sekä painettuna lehtenä että open access -julkaisuna. Sen osalta julkaisutietojen avoin saatavuus -tiedot eivät vaikuta kovin luotettavilta:

	määrä	osuus
Ei open access -julkaisu	42	37,8 %
Open access -julkaisu	11	10,0 %
Open access -julkaisu, rinnakkaistallennettu	5	4,5 %
Ei tietoa	53	47,8 %

Lehteä ei ole listattu DOAJ:hin, joten sen open access -status on jäänyt havaitsematta mm. Helsingin yliopistolta. Muuten saattaa olla, että lehden painettu versio on hämännyt raportoijia, koska vain kymmenessä prosentissa tapauksista on havaittu, että lehdellä on myös avoimessa käytössä oleva verkkoversio. Rinnakkaistallennetuiksi tulkituissa tapauksissa on tosin myös linkkejä lehden viralliseen verkkoversioon - sekaannus on ehkä ymmärrettävä siksi, että lehden verkkoversio ilmestyy nykyään THL:n julkaisuarkiston yhteydessä (<http://www.julkari.fi/handle/10024/90793>).

### 2.3. Rinnakkaistallennetut julkaisut

Rinnakkaistallennetuiksi ilmoitettuja julkaisuja on yliopistojen vuosien 2011-2013 datassa yhteensä 2854. Näistä noin 42 %:lle ei ole ilmoitettu minkäänlaista verkko-osoitetta (osalla tosin on DOI-tunniste). Suurella osasta muista (noin 24 %) on ilmoitettu ainoastaan DOI-pohjainen osoite, jonka voi olettaa vievän kustantajan palvelussa julkaistuu maksulliseen versioon.

Vain vajaalla 14 %:lla rinnakkaistallennetuiksi ilmoitetuista julkaisuista on validilta näyttävä URN-pohjainen osoite - käytännössä nämä julkaisut ovat pääosin peräisin neljästä yliopistosta (Jyväskylän yliopisto, Tampereen teknillinen yliopisto, Lapin yliopisto, Tampereen yliopisto). Helsingin yliopisto ei sen sijaan ole raportoinut omien rinnakkaistallenteittensa verkko-osoitteita, eli Heldaan viittaavia Handle-osoitteita esiintyy ainoastaan satunnaisesti muiden yliopistojen datassa.

Rinnakkaistallennetuiksi ilmoitettujen julkaisujen tiedoissa on paljon ei-pysyviä linkkejä mm. kustantajien palveluihin, käyttörajattuihin palveluihin ja satunnaisille verkkosivuille. Monissa tapauksissa vaikuttaa siltä, että kyse ei ole varsinaisista rinnakkaistallennetuista julkaisuista, eli joko kategorian merkitys on ymmärretty väärin tai julkaisu on sijoitettu vahingossa väärään kategoriaan. Koska datassa on vähän linkkejä varsinaisiin rinnakkaistallenteisiin, julkaisujen statusta on useimmissa tapauksissa mahdotonta verifioida.

OKM:n tiedonkeruussa saadut tiedot poikkeavat huomattavasti esim. suoraan yliopistojen julkaisuarkistoista kerätyistä rinnakkaistallennettujen julkaisujen määristä.<sup>3</sup> Poikkeamat eivät ole säännönmukaisia: Helsingin ja Jyväskylän yliopistojen osalta eri lähteistä saatujen julkaisumäärien suuruusluokat vastaavat jossain määrin toisiaan, mutta esim. Tampereen yliopistossa tiedonkeruuseen on päätyneet vain osa yliopiston Tampub-julkaisuarkistoon tallennetuista julkaisuista. Toisaalta Itä-Suomen yliopisto näyttää tiedonkeruun perusteella olleen yksi aktiivisimmista rinnakkaistallentajista (556 rinnakkaistallennettua julkaisua vuosina 2011-2013),

<sup>3</sup> Rinnakkaistallennusta koskevia tietoja on kerätty julkaisuarkistoista lähiaikoina julkaistavaa ATT-hankkeen julkaisutyöryhmän taustaselvitystä varten.

mutta tämä ei ole lainkaan todennettavissa yliopiston omasta julkaisuarkistosta, eikä oikeastaan julkaisudatastakaan, sillä joitakin poikkeuksia lukuun ottamatta Itä-Suomen yliopiston data ei sisällä linkkejä julkaisujen rinnakkaistallennettuihin versioihin. Muutamissa tapauksissa kyse saattaa toki olla muista open access -julkaisuista, jotka on tiedonkeruussa tällä hetkellä yhdistetty samaan kategoriaan.

Yleisvaikutelmaksi tiedonkeruun rinnakkaistallennusta koskevista tiedoista jää, että ne ovat toistaiseksi sekavia, sattumanvaraisia ja suurelta osin epäluotettavia.

### 3. Muita avoimen saatavuuden mittaamiseen liittyviä kysymyksiä

#### 3.1. Hybridijulkaisukanavat

- Varsinaisten open access -julkaisukanavien rinnalla etenkin suuret kansainväliset kustantajat ovat muuttaneet suuren osan julkaisukanavistaan ns. hybridilehdiksi, joissa yksittäiset artikkelit voi ostaa vapaaseen käyttöön maksua vastaan.
- Open access -liikkeen piirissä mallia pidetään ongelmallisena, eikä se ole tähän asti ollut kovin suosittu, mutta kustantajat markkinoivat sitä aktiivisesti.
- Tiedonkeruun ohjeiden perusteella hybridijulkaisukanavien OA-artikkelit sijoittuvat tällä hetkellä ilmeisesti samaan kategoriaan rinnakkaistallennettujen julkaisujen kanssa. Tämä on kuitenkin epäjohdonmukaista, koska tässä tapauksessa avoin saatavuus toteutuu kustantajan tarjoaman infrastruktuurin, ei avoimen julkaisuarkiston kautta.
- Kirjoittajamaksuihin perustuvan avoimen saatavuuden lisäksi maksulliset kanavat saattavat julkaista myös näyteluonteisesti osan artikkeleistaan OA:na (esim. tietyt teemanumerot tai yksittäisiä artikkeleita kustakin numerosta).

#### 3.2. Rinnakkaistallennuksen embargot

- Suuri osa rinnakkaistallennuksen sallivista julkaisukanavista on määritellyt artikkeleille julkaisuviiveen (embargon), jonka maksimipituudeksi suositellaan yleensä puolta vuotta (luonnontieteet) tai vuotta (humanistiset tieteet). Joillakin kanavilla on pidempiä embargoja.
- Embargot ovat jonkinasteinen haaste tietojen raportoinnin ja tarkistamisen kannalta: raportoidaanko embargon alaiset julkaisut rinnakkaistallennettuina jo ennen kuin ne ovat vapaasti saatavilla vai vasta embargon laukeamisen jälkeen? Tiedonkeruun kannalta voi olla perusteltua raportoida rinnakkaistallennus ja rinnakkaistallennetun version osoite mahdollisuuksien mukaan heti, vaikka julkaisu tulee saataville vasta myöhemmin. Toisaalta VIRT-mallisessa tiedonkeruussa saatavuutta koskevat tiedot voivat päivittyä myös myöhemmin.
- Embargo on julkaisukanavan ominaisuus. Lehtikohtaisia embargoja koskevia tietoja on mahdollista saada myös Sherpa/Romeon tietokannasta (<http://www.sherpa.ac.uk/romeo/>), joten embargojen pituutta koskevan tiedon kerääminen erikseen kunkin julkaisun raportoinnin yhteydessä ei välttämättä ole mielekästä.

#### 3.3. Viivästetty open access

- Osa lehdistä julkaisee artikkelinsa vapaasti verkossa esim. puolen vuoden tai vuoden viiveellä ("delayed OA") kunkin numeron ilmestymisestä. Näitä lehtiä ei tästä huolimatta yleensä lasketa OA-lehdiksi, eikä niitä hyväksytä mukaan esim. DOAJ:in listalle.

- Vastaavat embargot ovat rinnakkaistallennuksessa hyväksytyjä, eli OA:ta koskevat linjaukset ja tilastointikäytännöt ovat tältä osin epä johdonmukaisia myös kansainvälisellä tasolla.
- Viivästetty OA on joka tapauksessa julkaisukanavan ominaisuus, eli jos tällä periaatteella toimivista julkaisukanavista on kattava lista, sitä koskevia tietoja on mahdollista päivittää keskitetysti myös jälkikäteen.

### 3.4. Julkaisun versiot

- Kustantajilla ja julkaisukanavilla on vaihtelevia ehtoja sille, minkä version julkaisuista saa rinnakkaistallentaa. Osa kustantajista sallii lopullisen, taitetun kustantajan version tallentamisen, osa taas ainoastaan tutkijan viimeisen ”oman” (vertaisarvioidun) version tallentamisen.
- Lisäksi julkaisuja tallennetaan pre print -versioina. Esim. tiettyjen fysiikan alojen julkaisut tallennetaan hyvinkin kattavasti pre printteinä ArXiv-julkaisuarkistoon (<http://arxiv.org>), ja koska niiden viralliset vertaisarvioidut versiot ilmestyvät vasta myöhemmin, tutkijat käyttävät suurimmaksi osaksi nimenomaan näitä versioita.
- Julkaisuista saattaa olla liikkeellä useita erilaisia versioita – miten tärkeänä versioita koskevan tiedon keräämistä pidetään?
- Tiedonkeruun kannalta versioita koskevien tietojen kerääminen vaatisi joko uutta erillistä kenttää tai tietojen täydentämistä julkaisuarkistoista (olettaen että tieto on tallennettu standardimuodossa julkaisuarkistojen tietueisiin).

### 3.5. Hyväksytyt tallennuspaikat

- Kustantajan verkkopalvelun ja julkaisuarkiston lisäksi julkaisuja voi olla vapaasti saatavilla myös muualla, esim. tutkijan tai tutkimushankkeen omalla kotisivulla tai erilaisissa sosiaalisen median palveluissa. Tiedonkeruussa on kuitenkin syytä rajata sitä millaisissa palveluissa julkaisujen pitää olla saatavilla, jotta ne voi ilmoittaa OA-julkaisuina tai rinnakkaistallennettuina julkaisuina.
- Julkaisun pitää olla vapaasti saatavilla sellaisessa paikassa, joka takaa sen pysyvän saatavuuden. Julkaisun osoite pitää pystyä ilmoittamaan pysyviin tunnisteisiin (DOI, URN tai Handle) perustuvana pysyvänä osoitteena, joka resoloituu itse julkaisuun.
- Tutkijan tai tutkimushankkeen kotisivut eivät usein ole pitkäikäisiä, eli julkaisujen pysyvyydestä ei ole takeita.
- Useat kaupallisista lähtökohdista toimivat sosiaaliset verkkopalvelut (esim. Academia.edu, ResearchGate, Mendeley) ovat tutkijoiden keskuudessa suosittuja, mutta niiden pitkän aikavälin toimintaperiaatteista tai bisnesmalleista ei ole tietoa, eikä julkaisujen pysyvistä saatavuudesta ole takeita.
- Hyväksytyt tallennuspaikat on perusteltua ainakin tässä vaiheessa rajata kustantajien verkkopalveluihin, tutkimusorganisaatioiden omiin julkaisuarkistoihin ja tieteenalakohtaisiin julkaisuarkistoihin (mm. PubMed, ArXiv).

## 4. Julkaisujen verkko-osoitteiden raportointi

Kuten edellä on käynyt ilmi, yliopistoilta ja ammattikorkeakouluilta kerätyssä julkaisudatassa ei monissa tapauksissa ole mukana tyydyttäviä tietoja avointen julkaisujen verkko-osoitteista, minkä vuoksi julkaisujen avoimuuden todentaminen on monissa tapauksissa mahdotonta. Joissakin tapauksissa kyse on siitä, että tietoja ei ole kerätty paikalliseen tutkimustietojärjestelmään tai ne

ovat virheellisiä, joissakin taas siitä ettei paikalliseen järjestelmään kerättyjä tietoja välitetä eteenpäin OKM:n tiedonkeruuseen. Sekä julkaisutietojen verifiointiin että tiedonhakijoiden tarpeiden kannalta olisi erittäin toivottavaa, että julkaisujen verkko-osoitteet saataisiin kerättyä etenkin avoimesti saatavilla olevien julkaisujen osalta mahdollisimman kattavasti.

Tällä hetkellä julkaisujen verkko-osoitteet ilmoitetaan pääsääntöisesti kentässä ”pysyvä verkko-osoite”, jonka kuvaus on seuraava (Yliopistojen tiedonkeruun käsikirja 2014, s. 33):

**Pysyvä verkko-osoite**

Julkaisun pysyviin tunnisteisiin (esim. DOI, URN tai handle) pohjautuva verkko-osoite, joka vie suoraan julkaisun kokotekstiversioon (ei siis esim. lehden/kustantajan etusivulle). Voi viitata sekä vapaasti saatavilla olevaan versioon että kustantajan palvelimella sijaitsevaan käyttöoikeudeltaan rajoitettuun versioon.

**Tiedon muoto**

Verkko-osoite, esim. <http://dx.doi.org/10.1038/ng1104-1133>.

Lisäksi erilliseen kenttään ”DOI-tunniste” tallennettavasta DOI:sta on helppoa muodostaa toimiva DOI-pohjainen verkko-osoite, joka tai resolvoituu suoraan itse julkaisuun tai sen esittelysivulle.

Tavoite kerätä nimenomaan julkaisujen pysyviä osoitteita on sinällään perusteltu, sillä muunlaiset verkko-osoitteet lakkaavat usein toimimasta hyvinkin nopeasti, ja niiden päivittämisessä - olettaen että julkaisut ovat ylipäänsä enää saatavilla jossakin - olisi paljon työtä. Lisäksi tieteellisessä julkaisemisessa on erityisen perusteltua pyrkiä siihen, että viittauksien kohteena olevat julkaisut ovat saatavilla mahdollisimman pysyvästi. Sama koskee luonnollisesti myös julkaisujen avointa saatavuutta, eli julkaisujen satunnaista open access -saatavuutta esim. tutkijoiden tai tutkimushankkeiden kotisivuilla ei ole välttämättä edes mielekästä mitata.

Toisaalta osoitteiden pysyvyyden tiukka korostaminen on myös jossain määrin ongelmallista, sillä kaikilla julkaisuilla ei yksinkertaisesti ole pysyviä osoitteita, vaikka ne ovat verkossa saatavilla. Tämä koskee erityisesti ei-tieteellisiä julkaisuja, joiden osuus kaikista raportoiduista julkaisuista on merkittävä sekä yliopistoissa että etenkin ammattikorkeakouluissa. Tieteellisten julkaisukanavien osalta on mahdollista pyrkiä siihen, että ne ottavat pysyvät tunnisteet käyttöön, mutta populaarimpien kanavien motivoiminen tähän voi olla vaikeaa. Pysyvät osoitteet ovat siis tärkeä tavoite, mutta joiltakin osin voi olla mielekästä hyväksyä muunkinlaisia tietoja mukaan tiedonkeruuseen, etenkin jos halutaan tarjota tiedonhakijoille pääsy Juulin kautta näihin aineistoihin.

Toinen haaste on se, että yliopistojen järjestelmät saattavat tukea kunnolla ainoastaan DOI-tunnisteiden tallentamista. Esim. Helsingin yliopiston Elsevierin Pure-ohjelmistoon perustuvassa TUHAT:issa on oma kenttä DOI-tunnisteiden tallentamiseen, mutta ei muille pysyville tunnisteille. Muihinkin pysyviin tunnisteisiin perustuvia osoitteita voi tosin tallentaa erilliseen julkaisuun liittyville linkeille tarkoitettuun kenttään. Tähän kenttään tallennettuja tietoja ei kuitenkaan välitetä eteenpäin OKM:n tiedonkeruuseen, mahdollisesti siksi että se saattaa sisältää muitakin kuin pysyviä osoitteita. Käytännössä Helsingin yliopisto raportoi siis tiedonkeruuseen ainoastaan DOI-tunnisteita ja niihin pohjautuvia pysyviä osoitteita.

Yliopistojen julkaisudatassa rinnakkaistallennetuiksi merkityille julkaisuille on ilmoitettu vain harvoin erillinen nimenomaan rinnakkaistallennettuun versioon vievä verkko-osoite. Sen sijaan niille on kyllä



usein raportoitu DOI tai siihen pohjautuva verkko-osoite, joka vie kustantajan palvelimella sijaitsevaan maksulliseen versioon. Tältä osin olisi perusteltua tarkistaa tiedonkeruun käytäntöjä siten, että DOI:n ja siihen pohjautuvan osoitteen lisäksi myös rinnakkaistallennetun version pysyvä verkko-osoite saadaan kerättyä talteen.

## 5. Toimenpide-ehdotuksia

Avointa saatavuutta koskevan ohjeistuksen kehittämisessä on syytä pyrkiä siihen, että tietojen keruu perustuu mahdollisimman yksinkertaiseen ja selkeään kategorisointiin. Lisäksi olisi hyvä, että julkaisujen OA-statusta koskevien tietojen tallentaminen ja keruu olisi automatisoitavissa niin pitkälle kuin mahdollista, eikä tietojen täydentäminen aiheuttaisi raportoiville organisaatioille kohtuutonta työmäärää. Seuraavat toimenpiteet edistäisivät näiden päämäärien toteutumista:

1. Luovutaan erillisistä ”Ei open access -julkaisu” ja ”Ei tietoa”-kategorioista ja yhdistetään ne yhdeksi ”Ei vastausta”-kategoriaksi (arvo ”0”).
  - ”Ei tietoa”-vastausten suuri osuus ja kategoriaan liittyvät erilaiset käytännöt raportoivissa organisaatioissa tekevät tähänastisen avointa saatavuutta koskevan datan lähes käyttökelvottomaksi.
  - Avoin saatavuus -tiedon keruussa ensisijaisena tavoitteena pitäisi olla nimenomaan open access -julkaisujen tunnistaminen. Sen selvittäminen, ettei jokin julkaisu varmasti ole open access -julkaisu on nurinkurista ja todennäköisesti turhaa työtä - etenkin koska julkaisun status saattaa joka tapauksessa milloin tahansa muuttua, jos se rinnakkaistallennetaan.
  - Open access -julkaisujen osuus saattaa jäädä jossain määrin todellista alhaisemmaksi, mutta jos raportoivia organisaatioista tulevaisuudessa palkitaan open access -julkaisemisesta, niillä on selkeä intressi tunnistaa tuottamansa open access -julkaisut mahdollisimman kattavasti.
2. Open access -julkaisuille (etenkin rinnakkaistallennetut ja hybridijulkaisut) pitää raportoida avoimeen versioon johtava verkko-osoite.
  - Julkaisujen open access -statuksen verifiointi on usein mahdotonta ilman OA-versioon johtavaa linkkiä (lisäksi linkki on tiedonhakijoiden kannalta hyödyllinen).
  - Tieteellisissä OA-julkaisukanavissa ilmestyneiden julkaisujen osalta linkki ei ole periaatteessa aivan yhtä välttämätön (koska niiden OA-statuksen voi yleensä selvittää muutenkin), mutta kuitenkin toivottava.
  - Rinnakkaistallennetuille julkaisuille pitää ilmoittaa kustantajan palveluun vievän DOI-tunnisteen/osoitteen lisäksi myös linkki rinnakkaistallennettuun versioon.
  - Voi olla perusteltua pohtia sitä, tarvitaanko rinnakkaistallennetun version osoitteelle selvyiden vuoksi oma kenttä, vai onko järkevämpää, että kentässä ”pysyvä verkko-osoite” voidaan raportoida samalla kertaa useita osoitteita (tai riittääkö DOI:n osalta, että se on ilmoitettu tunnisteena omissa kentässään?).
  - Linkitykseen käytettävien osoitteiden on suositeltavaa perustua pysyviin tunnisteisiin (DOI, URN, Handle), mutta etenkin ei-tieteellisten julkaisujen kohdalla tästä voidaan joutua tinkimään - kuten käytännössä tähänkin asti on tehty.
3. Tarkistetaan tiedonkeruun ohjeistusta siten, että julkaisu voi olla samalla kertaa sekä OA-julkaisu (”1”) että rinnakkaistallennettu (”2”).
  - Open access -julkaisujen rinnakkaistallentaminen julkaisuarkistoon on melko yleistä.

- Toinen vaihtoehto on määritellä, kumman tyyppin avointa saatavuutta pidetään ensisijaisena tavoitteena.
4. Selvitetään, olisiko hybridilehdissä ilmestyneet OA-julkaisut ja muut ei-OA-julkaisukanavissa ilmestyneet OA-julkaisut mielekästä ilmoittaa mieluummin kategoriassa "Open access -julkaisu" kuin kategoriassa "Open access -julkaisu, rinnakkaistallennettu".
    - Hybridijulkaisuja on ainakin toistaiseksi niin vähän, ettei niille ehkä ole mielekästä luoda omaa kategoriaa.
    - Toisaalta ne sopivat huonosti rinnakkaistallennettujen julkaisujen yhteyteen, koska niiden tapauksessa avoin saatavuus toteutuu kustantajan infrastruktuurin, ei julkaisuarkistojen kautta.
  5. Kehitetään open access -julkaisujen tunnistamista tukevia kansallisia ja kansainvälisiä tietolähteitä ja niiden hyödyntämiseen liittyviä prosesseja.
    - Selvitetään, missä määrin eri tietolähteitä on mahdollista hyödyntää paikallisella tasolla kunkin raportoivan organisaation prosesseissa ja missä määrin niiden avulla on järkevää täydentää/rikastaa organisaatioilta kerättäviä tietoja keskitetysti tiedonkeruun yhteydessä/Juulissa.
    - Selvitetään mahdollisuuksia haravoida julkaisuarkistoista automaattisesti rinnakkaistallennettujen julkaisujen tietoja ja muita keinoja kytkeä julkaisuarkistot nykyistä paremmin mukaan tiedonkeruun prosesseihin.
    - Kannustetaan kotimaisia open access -lehtiä liittymään mukaan DOAJ:hin.
    - Selvitetään, riittääkö DOAJ:n lista tiedonkeruun tarpeisiin, vai olisiko Julkaisufoorumin lehtilistan yhteydessä mielekästä pitää yllä myös muuta tietoa julkaisukanavien open access -statuksesta (ml. tiedot viivästetystä OA:sta).
    - Ylläpidetään tietoja julkaisukanavien Sherpa/Romeo-statuksesta ja tuotetaan sen pohjalta tietoja rinnakkaistallennuksen potentiaalisesta määrästä suhteessa käytännössä toteutuneeseen rinnakkaistallennukseen.