

2015-11-04

Julkaisuarkistopalveluiden asiantuntijaryhmä, 4.11.2015

- [Julkaisuarkistopalveluiden asiantuntijaryhmä, 4.11.2015](#)
 - [1. Kokouksen avaus ja puheenjohtajan valinta](#)
 - [2. OpenAire-yhteensopivuus](#)
 - [3. Julkaisuarkistoaineistot ja Finna](#)
 - [4. Julkaisuarkistopalveluihin liittyvät koulutukset](#)
 - [5. Muut asiat](#)
 - [6. Seuraava kokous](#)

Aika: Keskiviikko 4.11.2015 klo 10.00-12.00

Paikka: Yliopistonkatu 1, Fabianian auditorio, 2.krs, Helsinki

Puheenjohtaja: Mikko Pennanen (Turun yliopisto)

Sihteeri: Samu Viita (Kansalliskirjasto)

Läsnä: Minna Marjamaa (AMKIT konsortio), Iris Tahvanainen (Lappeenrannan teknillinen yliopisto), Minna Liikala (THL), Jukka Lehmus (Kansalliskirjasto), Jyrki Ilva (Kansalliskirjasto), Päivi Rosenström (Kansalliskirjasto), Samu Viita (Kansalliskirjasto)

Etänä: Mikko Pennanen (Turun yliopisto), Tiina Tolonen (AMKIT-konsortio), Mikko Saunamäki (Luonnonvarakeskus)

Poissa: Teemu Makkonen (Jamk), Nina Hyvönen (Kansalliskirjasto), Toni Raja-Hanhela (Lapin yliopisto)

1. Kokouksen avaus ja puheenjohtajan valinta

- Valittiin työryhmän puheenjohtajaksi Mikko Pennanen Turun yliopiston kirjastosta. Puheenjohtaja toimii tehtävässä toistaiseksi, ja hän vastaa kokousten asialistan koostamisesta yhdessä Kansalliskirjaston asiantuntijoiden kanssa. Ryhmän jäsenten toiveet huomioidaan asialistan suunnittelussa.

2. OpenAire-yhteensopivuus

- Samu Viita esitteli Lappeenrannan yliopiston kanssa toteutetun OpenAire-pilotin tähänastisia tuloksia.
- Pilotin valmistelu aloitettiin poimimalla OpenAiren spesifikaatioista pakolliset kentät, minkä jälkeen LutPubin artikkelien syöttölomaketta muutettiin vastaamaan OpenAiren vaatimusten mukaiseksi. LutPubille toteutettiin OAI2-rajapinnan avulla oma setti, johon valikoituu vain Lappeenrannan teknillisen yliopiston tietueita ja joissa on OpenAiren spesifikaation mukaiset kentät ja arvot käytössä. Piloti on lähes valmis ja on nyt vielä ohjeistuksen ja syöttöjärjestyksen osalta LutPubin täydennettävänä. Tämän jälkeen voidaan aloittaa tuotantokäyttö. Huolena LutPubissa osalta on tosin se, tuleeko Converis-tutkimustietojärjestelmän käytöstä jatkossa lisähaasteita OpenAire-yhteensopivuudelle. Converiksesta on tarkoitus haravoida tietueita Doriaan ja tähän käytetään Converiksen Pentaho-rajapintaa, joka antaa vastauksena OAI:ta. Todettiin, että OpenAiren yhteensopivuusasiaan saadaan selkeyttä kun Pentaho-rajapintaa päästään testaamaan ja tarkistamaan, saadaanko sitä kautta tarvittavat tiedot ja tarvittavassa muodossa.
- Todettiin, että pilotin myötä on varmistunut, että asiakasorganisaatioille on tarvittaessa tehtävissä omat OpenAire-setit myös Dorian kaltaisissa jaetuissa julkaisuarkistoissa. Näin organisaatio saadaan tarvittaessa näkymään itsenäisenä OpenAiressa sen sijaan, että aineistot näkyisivät siellä vain arkistokohtaisesti.
- Pilotin yhteydessä selvisi myös, että EU-projektikoodi ei ole pakollinen tieto, vaikka DSpaceen OpenAire-toteutuksessa se on alun perin asetettu pakolliseksi tiedoksi. Ainakin LutPub:issa haluttaisiin kuitenkin tämä pakolliseksi tiedoksi, jota ilman julkaisu ei näy OpenAire-setissä.
- Keskusteltiin muiden osallistuvien organisaatioiden liittymisestä OpenAiren tiedonkeruuseen. Esim. Turun yliopisto haluaisi liittyä OpenAireen, mutta avoimena on vielä ainakin se, miten Converiksen ja Dorian tiedot sovitetaan yhteen OpenAiren haravoinnissa. THL:n osalta puolestaan pohdittiin, miten OKM:n tiedonkeruun kuvailutiedoille asettamat vaatimukset saadaan sovitettua OpenAiren vaatimuksiin. Jukurin osalta haasteena ovat organisaation sisäiset projektikoodit, joille löytyy tosin vastineensa EU-projektikoodista. Jos Jukuriin toteutetaan OpenAire-yhteensopivuus, täytyy selvittää, olisiko määppäys hoidettavissa koodin avulla tai esim. massaeditorilla. Theseuksen osalta todettiin, että sen kuvailutiedot olisi hyvä saada näkymään OpenAiressa oikein. Tällä hetkellä Theseus näkyy kokonaisuudessaan OpenAiressa, vaikka sen metadata ei olekaan OpenAire-yhteensopivaa.
- Pohdittiin myös, että minkä aineistojen olisi syytä näkyä OpenAire-setissä. OpenAiren spesifikaation eri versiot ovat tältä osin ristiriitaisia. Aikaisemmin OpenAire keskittyi EU-projektikoodilla varustettuun aineistoon, mutta uudempien versioiden kriteerit ovat löyhempiä. Kun asia tulee Theseuksen osalta ajankohtaiseksi, täytyy selvittää vielä lisää sitä, ovatko opinnäytteet todella OpenAiren kannalta relevanttia aineistoa.

3. Julkaisuarkistoaineistot ja Finna

- Finna-palvelu tarjoaa suomalaisten arkistojen, kirjastojen ja museoiden aineistoja yhdellä haulla. Jukka Lehmus oli paikalla kokouksessa teknisenä asiantuntijana vastaamassa Finnaa koskeviin kysymyksiin.
- Jyrki Ilva piti alustuksen aiheesta, nostaen esiin kysymyksen siitä, mistä lähteistä Finnaan on ensisijaisesti tarkoitus tuoda avoimen tieteen aineistojen kuvailutietoja. Julkaisutietoja on saatavilla monesta lähteestä, kuten tutkimustietojärjestelmä, julkaisuarkistot, kirjastotietokannat ja Melinda. Julkaisutietoja viedään myös OKM:n tiedonkeruun kautta CSC:n ylläpitämään Virta-tietovarantoon, jonka pitkän tähtäimen tavoitteena on olla kansallinen tutkimustietojärjestelmä. CSC on automatisoimassa vuosien 2015 ja 2016 aikana tietojen keruuta palveluun ja kehittää rajapintoja, joiden kautta kerätyt tiedot ovat haravoitavissa edelleen muihin järjestelmiin. Virta-tietovarannon sisältämät tiedot ovat selattavissa Kansalliskirjaston kehittämän Juuli-julkaisutietoportaalin kautta. Juuli on julkaisutietojen selailuun ja tiedonhakuun räätälöity käyttöliittymä, joka perustuu Finnan tavoin VuFind-ohjelmistoon. Julkaisujen avointa saatavuutta olisi myös jatkossa mahdollista monitoroida kansallisella tasolla sen kautta.
- Finnan osalta haasteena on mm. se, mistä palvelusta kukin organisaatio toimittaa tietoaan Finnaan. Tällä hetkellä etenkin kansallisessa näkyvässä ei ole juuri lainkaan julkaisuarkistoista haravoituja kuvailutietoja. Esim. Finna.fi:stä löytyvien opinnäytteiden tiedot ovat peräisin kirjastotietokannoista. Jyrki esitti lopuksi myös kaaviokuvan, jossa oli hahmoteltu nykyisiä ja mahdollisia tulevia tietovirtoja eri järjestelmien välillä.
- Keskusteltiin Finnan moninaisuudesta ja sen aiheuttamista haasteista selkeyden ja käytettävyyden suhteen: Kansallisen Finnan lisäksi monilla yliopistoilla ja amk:eilla on organisaatiokohtainen Finna, joka sisältää pääasiassa organisaation "omia" aineistoja. Esim. amk:eissa kansallista Finnaa käytetään suhteellisen vähän, koska korkeakoulut haluavat koostaa omaa aineistoaan opiskelijoille yms. kohderyhmille. Loppukäyttäjän kannalta eri Finnojen olemassaolo voi kuitenkin aiheuttaa hämmennystä, koska Finnaa on usein markkinoitu kansallisen Finnan näkökulmasta, eli palveluna joka tarjoaa kaiken tiedon yhdestä osoitteesta.
- Luke kysyi, olisiko jotain vakioitua tapaa viedä aineistoa Finnaan sitten kun siihen liittyy. Jukka Lehmus totesi, että ainakaan tällä hetkellä vakioitua tapaa ei ole, vaan aineistoja tulee useammasta lähteestä ja tapauskohtaisesti.
- Tieteellisten artikkelien osalta nousi esiin kysymys siitä, olisiko niiden kuvailutiedot mielekästä viedä Primo Central Indexiin (PCI), jolloin ne olisivat haettavissa samalla kertaa vastaavien kv. aineistojen kanssa. PCI koetaan kuitenkin ongelmalliseksi, koska kyseessä on yhden palveluntarjoajan maksullinen tuote, joka saattaa jossain vaiheessa kilpailutusten yhteydessä vaihtua johonkin toiseen tuotteeseen.
- Jotta Finnan kansallisen näkymän metadata voitaisiin antaa jatkokäyttöön avoimena datana, sen edellytetään olevan CC0-lisensoitua eli tekijänoikeudesta vapaata. Yllättäen on ilmennyt, että osa korkeakouluista kokee tämän ongelmallisena etenkin opinnäytetöihin liittyen. Huoli liittyy kuvailutietojen sisältämiin henkilötietoihin, joiden CC0-lisensoinnin pelätään mahdollistavan ulkopuolisille tahoille korkeakoulujen opiskelijoita koskevien henkilörekisterien muodostamisen. Keskustelu on lähtenyt alun perin liikkeelle Aalto-yliopistossa tietosuojavaltuutetun lausunnoista tehdyistä tulkinnoista ja se on noussut aiemmin esiin mm. kirjastoverkkopalveluiden keskustelupalstalla. Nyt myös Lappeenrannan teknillisessä yliopistossa on lakimiehen suosituksesta päätetty, ettei opinnäytteiden tietoja toistaiseksi viedä kansalliseen Finnaan, joten asiaan liittyvästä epäselvyydestä on tullut merkittävä hidaste Finnan kattavuuden laajenemiselle etenkin opinnäytteiden osalta.
- Samaa asiaa on pohdittu Theseuksen osalta Finnaan viennin yhteydessä ja sen suhteen on todettu, että ongelmaa ei ole, koska AMKIT-konsortio omistaa kuvailutiedot. Kuvailutiedot eivät tämän selvityksen mukaan ole kuvailijan, esimerkiksi opiskelijan omaisuutta, vaan kuuluvat järjestelmän omistajalle, joka tietoja kerää. Opiskelija voi pidättä kaikki oikeudet opinnäytteeseensä tai antaa sille CC-lisenssin, mutta oikeudet kohdistuvat vain itse teokseen, eli tässä tapauksessa opinnäytetyöhön. Kuvailutietojen katsotaan kuitenkin Theseuksen osalta kuuluvan järjestelmän omistajalle. Theseuksen yhteydessä tekemä selvitys ei kuitenkaan riitä kattamaan kaikkia muita Finnan lähteitä ja taustajärjestelmiä.
- Keskustelussa todettiin, että opinnäytteiden kuvailutiedot ovat käytännössä jo nyt suurelta osin vapaasti saatavilla ja niitä haravoidaan mm. julkaisuarkistoista monenlaisiin muihin palveluihin. Todettiin myös, että CC0-lisensointi koskee aineistoon liittyviä tekijänoikeuksia, eikä sillä pitäisi sinällään olla mitään kytkeä henkilötietoihin liittyviin tietosuojakysymyksiin. Koska asia kuitenkin koetaan epäselväksi, toivottiin että sitä selvitettäisiin myös kansallisella tasolla. Jyrki Ilva ehdotti, että asian voisi viedä esim. ATT-hankkeen oikeuksien hallinta -työryhmän pohdittavaksi.
- Julkaisuarkistoaineistojen osalta on pohdittu myös sitä, miten erilaiset käyttöoikeusrajoitukset voidaan ilmaista Finnalle, jotta Finnan käyttöliittymässä voidaan mainita asiasta. Tämä on syytä toteuttaa julkaisuarkistopuolella huolella ja metadataan pohjautuen. Haasteita on siinä, että rajoituksia voi olla monenlaisia, kuten että tietueen tiedostot on IP-rajoitettu, käyttäjätunnistukseen perustuen rajoitettu tai julkaisuarkiston kuvailutiedoissa oleva tiedostolinkki tai linkit osoittavat toiselle palvelimelle, joista osa tai kaikki ovat IP-rajoitettuja. Lisäksi embargo pitäisi ilmoittaa vakioidulla tavalla. Tarvittaisiin siis yleiset säännöt, miten Finnalle ilmoitetaan asia metadatan avulla. Julkaisuarkistojen osalta asia nousi esiin, kun Sibelius-Akatemian liityessä Finnaan tuli tarve ilmaista asia Finnalle räätälöidyssä OAI-metadataformaattissa (qdc_finna). Jukka Lehmus totesi, että tämä on haasteellinen ongelma ja ratkaisua pyritään löytämään ensin MARC-aineistoille ja vasta sitten Dublin Coren osalta, joten ongelman ratkaisemiseen voi mennä vielä jonkin verran aikaa. Finnan tavoite on myös saada käyttöoikeusrajoituksia koskevat tiedot koneluettavaan muotoon.

4. Julkaisuarkistopalveluihin liittyvät koulutukset

- Aikataulullisista syistä johtuen koulutusta ei pystytty järjestämään samana päivänä kuin tämä kokous, mutta se pyritään järjestämään loppuvuoden aikana. Päivämäärästä laaditaan kysely kohderyhmälle ja valitaan aika sen perusteella.

5. Muut asiat

- Uusia organisaatioita tulossa mukaan palveluiden asiakkaina
 - Ahvenanmaan ammattikorkeakoulu aloittaa ensi vuoden alusta Theseuksessa
 - Eläketurvakeskus on liittymässä joulukuussa Julkarin asiakkaaksi
 - Valtioneuvoston kanslialle luodaan uusi instanssi, johon siirretään Julkarista STM:n aineistot. Instanssiin tulevat kaikki ministeriöt, mutta ei niiden alaisia laitoksia.
- Helsingin yliopiston muutosohjelman vaikutukset palveluiden toimintaan
 - Rekrytointikielto vaikeuttaa tällä hetkellä mm. uusien hankkeiden käynnistämistä
 - Julkaisuarkistopalveluiden toiminta pyritään kuitenkin takaamaan

6. Seuraava kokous

- Kokous pidetään 27.1. klo 10-12, paikka ilmoitetaan myöhemmin. Kokouksen aihe-ehdotuksia voi lähettää kokouksen puheenjohtajalle Mikko Pennaselle osoitteeseen: mikko.piste.pennanen@miukumauku.fi