

2015-04-09

Julkaisuarkistojen asiantuntijaryhmä 9.4.2015

1. Ryhmän järjestäytyminen
2. Ryhmän tehtävät yleisellä tasolla
3. Ryhmän tehtävät vuonna 2015
 - 3.1 Palveluiden yleisesittely
 - 3.2 Kehityssuunnat kansainvälisessä DSpace-yhteisössä: mitä KK on ajatellut ottaa kehittämiseen mukaan?
 - 3.3 Avoin tiede ja tutkimus -hanke
 - 3.4 Julkaisarkistosopimusten päivittäminen
 - 3.5 ORCID
 - 3.6 CRIS-kytkennät
 - 3.7 Tilastointi
4. OpenAire -yhteensopivuus
5. Muut asiat

Aika: 9.4.2015 klo 9.00-11.00

Paikka: Teollisuuskatu 23, 4.krs, C454 (Tiuhti), Helsinki

Sihteeri: Päivi Rosenström

Läsnä: Minna Marjamaa (AMKIT konsortio), Iris Tahvanainen (Lappeenrannan teknillinen yliopisto), Mikko Pennanen (Turun yliopisto), Nina Hyvönen (Kansalliskirjasto), Jyrki Ilva (Kansalliskirjasto), Samu Viita (Kansalliskirjasto), Päivi Rosenström (Kansalliskirjasto)

Etänä: Teemu Makkonen (Jyväskylän Ammattikorkeakoulun kirjasto, poistui 10.30) ja Toni Raja-Hanhela (Lapin yliopisto)

Poissa: Mikko Saunamäki (Luonnonvarakeskus), Tiina Tolonen (AMKIT-konsortio) , Minna Liikala (THL)

1. Ryhmän järjestäytyminen

Työryhmän jäsenet esittäytyivät. Puheenjohtajaksi valittiin Nina Hyvönen ja sihteeriksi Päivi Rosenström. Kokouksen aluksi esityslistaa täydennettiin kohdilla Tilastointi ja CRIS-kytkennät. Todettiin, että ryhmää voidaan vielä täydentää ja sen kokouksiin voi osallistua muidenkin asiakasorganisaatioiden edustajia tarpeen mukaan. Tiedonhallinnan ohjausryhmän toivomus asiantuntijaryhmän kooksi oli 8-10 jäsentä. Pöytäkirjojen saatavuudesta tullaan huolehtimaan. Tämä on yhteydessä Kansalliskirjaston verkkosivu-uudistukseen. Pöytäkirjojen lopullinen sijoituspaikka on todennäköisesti kiwi-palvelu.

2. Ryhmän tehtävät yleisellä tasolla

Ryhmä on suunnattu ensisijaisesti Kansalliskirjaston julkaisuarkistoasiakkaille. Julkaisuarkistojen asiantuntijaryhmän tehtäviin kuuluu Kansalliskirjaston julkaisuarkistojen kehittämislinjojen suunnittelu. Julkaisuarkistojen asiantuntijaryhmä ei ole päättävä elin vaan se vie ehdotuksensa tarvittaessa tiedonhallinnan ohjausryhmän päätettäväksi. Tämä ei tarkoita sitä, että kaikki kehitysideat olisi pakko viedä virallisen päätöksentekoprosessin kautta. Mikäli ryhmässä syntyy hyviä ideoita, niitä otetaan toteutukseen suoraan, mikäli resurssit sen sallivat.

ATT-hankkeen puitteissa on tulossa enemmän julkaisuarkistoihin ja open accessiin liittyvää koordinoitua. On hyvin mahdollista, että asiantuntijaryhmän rinnalle perustetaan toinen yhteisiä käytäntöjä koordinoiva ryhmä, jossa olisi jäsenenä muitakin kuin Kansalliskirjaston palveluiden asiakkaita.

3. Ryhmän tehtävät vuonna 2015

3.1 Palveluiden yleisesittely

Jyrki Ilva [piti esityksen Kansalliskirjaston julkaisuarkistopalveluista](#). Tällä hetkellä palveluissa on mukana 40 asiakasorganisaatiota. Infrastruktuurissa on myös Kansalliskirjaston omia palveluita ja aineistoja. Työnjako asiakkaiden ja Kansalliskirjaston välillä on se, että Kansalliskirjasto vastaa arkistojen teknisestä ylläpidosta, varmuuskopioinnista, versiopäivityksistä ja asiakkaille annettavasta koulutuksesta. Asiakasorganisaatiot ovat vastuussa julkaisuarkistojensa sisällöstä.

Kansalliskirjaston palvelut koostuvat edellä mainitusta peruspalvelusta ja lisäksi on erillishintaan mahdollista saada lisäpalveluita kuten käyttöliittymien muokkausta, kytkentöjä muihin järjestelmiin, konversioita, eräajoja sekä oma DSpace-instanssi.

Kansalliskirjaston julkaisuarkistopalveluilla ei ole keskitettyä kansallista rahoitusta vaan kaikki rahoitus tulee asiakasorganisaatioilta. Toiminta on tasapainoilua yksittäisille asiakkaille tehtävän työn ja infrastruktuurin yleisen kehittämisen välillä. Käytännössä asiakasorganisaatioille tuotettavien palveluiden parissa työskentelee kokopäiväisesti kaksi henkilöä (Samu Viita ja Päivi Rosenström), mutta heidän lisäksi tekniseen kehittämiseen /ylläpitoon ja Kansalliskirjaston omiin aineistoihin liittyvään työhön osallistuu muitakin henkilöitä. Jyrki Ilva ja Nina Hyvönen hoitavat sopimus- ja asiakassuhteisiin liittyviä asioita. ATT-hanke saattaa muuttaa tilannetta. Sieltä saadaan ehkä yksi tekninen henkilö lisää kahdeksi vuodeksi.

Keskustelussa tuli esille, että kaivattaisiin yhteisiä pelisääntöjä aineistojen poistamiseen huomioiden mm. vapaakappaleasiat. Todettiin, että Kansalliskirjaston laatimille kansallisille suosituksille metadatakenttien käytön suhteen olisi käyttöä. Tämä vaikuttaa esimerkiksi OpenAire-yhteensopivuusasioihin. Opinnäytteiden suhteen tilanne on kuitenkin ollut hyvä kiitos aikanaan laaditu opinnäytteisiin kohdistuvan ohjeen. Lisähaastetta tuo se, että kansainvälisestikin Dublin Core -formaatin kehityksessä on ongelmia, jotka sitten heijastuvat myös Suomeen.

3.2 Kehityssuunnat kansainvälisessä DSpace-yhteisössä: mitä KK on ajatellut ottaa kehittämiseen mukaan?

Kansalliskirjasto seuraa aktiivisesti kansainvälistä DSpace-kehitystä, mutta päivittää uusiin versioihin viiveellä, jotta niistä on jo selvitetty mahdolliset ongelmat. REST-rajapinnan kehittämisen osalta Kansalliskirjasto oli esillä myös kv. tasolla. Kansalliskirjastossa kehitetty SimpleREST-rajapinta oli mukana toisena loppusuoralle yltäneistä vaihtoehdoista kun DSpaceen valittiin virallista REST-toteutusta. Se ei kuitenkaan tullut valituksi osittain aikataulullisista syistä johtuen.

Lähitulevaisuuden kehityskohteista Kansalliskirjastossa priorisoidaan mm. parannuksia varmistustekniikoihin (AIP-palautus), seuraavan versiopäivityksen suunnittelua, SYLI:n kehitystä, Crowdin käyttöönottoa ja SimpleREST:n kehitystyötä.

Varmistustekniikan parannuksessa on kyse siitä, että aiemmin mahdollinen palautus on jouduttu tekemään nauhalta, joka vaatii käytännössä koko arkiston palauttamisen tiettyyn tilaan. Nyt on suunnitteilla DSpaceen AIP-paketteihin perustuva tekniikka, joka mahdollistaa valmistuessaan yksittäisten yhteisöjen palauttamisen ilman että koko arkistoa pitää palauttaa ongelmaa edeltävään tilaan.

Seuraavan versiopäivityksen suunnittelu on myös alkamassa. Näillä näkymin seuraava käyttöönotettava versio on mahdollisesti DSpace 5. Aikataulusta on vaikea sanoa tässä vaiheessa vielä mitään varmaa, mutta todennäköisesti päivitys tapahtuu vuoden 2016 aikana. DSpace 5 tuo muutoksia käyttöliittymiin (Discovery-moduli ja mobiilikäyttöä nykyistä paremmin tukeva Mirage 2-käyttöliittymäteema). Discovery-modulin mukanaan tuomat Finnan käyttöliittymän kaltaiset fasetit asettavat vaatimuksia myös metadatalle. Se kunnolla toimii vain jos metadata on yhtenäistä. On mahdollista, että versio ja aikataulu muuttuu vielä: päivityksissä ilmenee usein yllätyksiä, jotka joudutaan korjaamaan. Siihen kuluu aikaa. DSpace 6 ilmestyy todennäköisesti vuodenvaihteen tienoilla.

SYLI:n ja SimpleREST:n kehitys ovat korkealla prioriteettilistalla. Åbo Akademilla on SYLI:in liittyvä pilotti menossa. Sen puitteissa on havaittu, että SYLI-järjestelmää on tarve keventää. SYLI:n kehitys etenee verkkaisesti, koska siihen ei ole ollut osoittaa täysipäiväistä henkilöä. SYLI-projektiin liittyy olennaisesti Crowd-järjestelmän käyttöönotto, joka mahdollistaisi sekä shibbolethin että LDAP:n käytön tunnistautumisessa. Crowdin käyttöönotto on ollut tauolla kiireellisempien tehtävien vuoksi. Crowdiin voidaan liittää Identity providereita, joiden avulla tunnistautuminen onnistuu. SYLI:ssä Crowd toimii REST-rajapinnan avulla. Crowdin käyttö on KVP:ssä yhteisesti sovittu asia. SYLI:n valmistuminen olisi tärkeää, koska sen avulla olisi mahdollista toteuttaa myös videoita sisältävien opinnäytteiden streamaus. Tämä olisi tärkeää AMKIT-konsortiolle.

SimpleREST:n kehitykseen on saatu lisäresursseja Metropoliassa käynnissä olevan opinnäyteprojektin avulla.

3.3 Avoin tiede ja tutkimus -hanke

Kansalliskirjasto on ollut aktiivisesti mukana useissa ATT:n työryhmissä. Kansalliskirjasto on lähiaikoina saamassa ATT:n puitteissa OKM:lta rahoitusta mm. julkaisuarkistoihin liittyvään hankekokonaisuuteen.

Kansalliskirjaston tulevan ATT-hankekokonaisuuden puitteissa pyritään yhdenmukaistamaan mm. julkaisuarkistoissa käytettyyn metadataan liittyviä käytäntöjä. Keskustelussa todettiin, että Dublin Corea koskeva kansallinen suositus olisi todella hyvä asia. Open Aire -yhteensopivuuden edellyttämät vaatimukset monimutkaistavat metadatakuvioita, joten suositukselle olisi tarvetta. Kehityssuunnista SYLI-projektilla on ykkösprioriteetti, mutta sen suhteen on muistettava että sitä tehdään muiden projektien rinnalla. Esimerkiksi uusien asiakkaiden käyttöönottoprojekti vievät aikaa. Doriasta sanottiin, että se on ollut vakaa järjestelmä ja sen eteen tehtyä työtä arvostetaan. Pohdittiin sitä, kuinka paljon tarvittaisiin lisärahoitusta, jotta kehitysprojekteja saataisiin toteutettua. Lähitulevaisuudessa on ehkä mahdollisuus saada lisähenkilö tekniseen kehittämiseen. Julkaisuarkistojen painoarvoa lisää tulevaisuudessa se, että rinnakaistallentamista koskevia mandaatteja tulee voimaan entistä useammissa organisaatioissa.

Toisaalta tutkijoille on välillä epäselvää, miksi artikkeleita tallennetaan Doriaan eikä esimerkiksi ResearchGate-palveluun. Dorian käyttöä puoltaa se, että ResearchGate ja vastaavat palvelut ovat kaupallisia palveluita. Ei ole mitään takeita siitä, ovatko ne olemassa vielä viiden vuoden päästä. Palveluissa ei ole myöskään käytössä pysyviä tunnisteita eikä aineistoa tallenneta pysyvästi. Tutkijat kokevat kuitenkin ResearchGaten kaltaiset palvelut hyödyllisinä, koska ne tarjoavat tallennuspalveluiden lisäksi verkostoitumismahdollisuuden.

3.4 Julkaisarkistosopimusten päivittäminen

SYN:n puolelta on noussut esille kysymyksiä sopimusten uusimisesta. Aikaisemmin kaikille oli räätälöity sopimusmalli. Sitten sopimuksia on pyritty yhtenäistämään siten, että kaikille on samanlainen sopimus pohja. Lisäliitteillä on sovittu käyttöönottoprojektien mm. aikatauluista ja lisätoista.

Keskusteltiin siitä onko muilla asiakkailla tarvetta muutoksille. AMKIT-konsortiolle on kiinnostusta käydä sopimusta läpi. LUT:ssa sopimusten tarkistaminen ei ole kiireellinen asia. Sovittiin, että asiaa katsotaan tämän vuoden aikana. Mahdollisesti jo seuraavassa kokouksessa käydään läpi sopimus pohjaa. Sovittiin, että sen saa etukäteen katsottavaksi.

Theseus-sopimusta voitaisiin käsitellä Theseus-ohjausryhmässä.

3.5 ORCID

Jyrki Ilva on ollut mukana kansallisissa tutkijan tunnistamista selvittäneissä työryhmissä. Asiaan liittyy jonkin verran lakitekniisiä kysymyksiä, mutta ORCID-tunnisteen kanssa on tarkoitus edetä. Auki on vielä tehdäänkö asiasta kansallinen konsortiosopimus vai solmitaanko yksittäisiä organisaatiokohtaisia sopimuksia, koska suuri osa etenkin pienemmistä tutkimusorganisaatioista pärjää todennäköisesti myös ORCID:in julkisella API:lla. Kustannusten jako OKM:n ja organisaatioiden kesken on myös edelleen auki.

Julkaisuarkistojen osalta kannattaa odottaa kansallista ratkaisua. Tunnisteita ei luoda keskitetysti kaikille vaan tutkijat luovat ne itse. ORCID-kytkentä DSpaceen on mahdollinen ja sitä on kehitelty maailmalla.

Keskusteltiin siitä, että rinnakaistallennuksen kohdalla ORCID-tunnisteiden käytössä voi olla ongelmia jos kirjasto syöttää tietoja tutkijoiden puolesta, koska tutkijoiden pitäisi periaatteessa tunnistautua itse ORCID:in järjestelmään. Kentällä on kuitenkin vahvana trendinä se, että tutkijoiden puolesta tehdään asioita, jotta he voivat keskittyä itse tutkimustyöhön. Akuutein haaste julkaisuarkistojen kannalta on se, miten ORCID saadaan ilmaistua mielekkäästi Dublin Coressa. Tässä voidaan ja kannattaa seurata kansainvälistä keskustelua. Hanna-Mari Puuska (CSC) on ollut keskeisessä roolissa selvityksen laatimisessa. Aihepiiristä järjestetään 23.4. kansallinen seminaari.

ISNI:n ja ORCID:n suhde on jossain määrin jännitteinen. ORCID-tunnisteet ovat joka tapauksessa muodoltaan yhdenmukaisia ISNI:n kanssa.

3.6 CRIS-kytkennät

Julkaisuarkistotyössä pitää tulevaisuudessa ottaa entistä enemmän huomioon yliopistojen rahoitusmalliin ja OKM:n tiedonkeruuseen liittyvät tarpeet. Open access -julkaisujen määrä on mahdollisesti tulossa kriteeriksi OKM:n rahoitusmalliin vuodesta 2019 alkaen. Avointa saatavuutta koskevien tietojen laatua pitää kuitenkin parantaa, ennen kuin niitä voidaan käyttää rahoituksen pohjana.

Aihe herätti seuraavaa keskustelua. Miten Dorialla pystytään vastaamaan tähän vaatimukseen? Lappeenrannalla on rehtorin määräämä mandaatti rinnakaistallentaa julkaisut Doriaan. Tätä varten tarvittaisiin joko shibboleth-kirjautumismahdollisuus Doriaan, jotta tutkijat voisivat itse viedä julkaisunsa Doriaan tai sitten mahdollisuus haravoida data Converiksesta Doriaan. Teknisesti shibboleth-kirjautuminen vaatisi joko oman instanssin tai sitten SYLI-syöttölomakkeen käyttöä. Shibboleth-kirjautuminen ei ole mahdollista suoraan Doriaan, koska se on usean organisaation julkaisuarkisto. SYLI-syöttölomakejärjestelmän saamisesta tuotantokäyttöön on toistaiseksi vaikea antaa luotettavaa arviota.

Converiksen suhteen haravointia on testattu ja se toimii ainakin metadatan osalta. Lappeenrannan teknillisen yliopiston kannalta olisi tärkeää selvittää, mitä moduuleja Converikseen vaaditaan, jotta haravointi onnistuu. Lappeenrannassa ollaan juuri päivittämässä Converista versioon 5 ja vaatimusmäärittelyt ovat käynnissä.

Iris Tahvanainen ehdotti, että Lappeenrannan yliopiston Converis-Doria-linkkaus voisi olla hyvä referenssitapaus DSpace-CRIS-linkkauksesta Kansalliskirjastolle. Lappeenrannan puolesta Kari Paakkunainen ottaa yhteyttä asian suhteen ja lisätietoja voidaan pyytää Åbo Akademista, jonka Converiksesta on testattu hakea metadatta Dspaceen. Selvitettäväksi jää onko Converiksesta mahdollista siirtää myös tiedostot rajapinnan kautta. OAI-ORE-tuen avulla tämä voisi olla mahdollista.

Todettiin, että Thomson-Reutersin Converiksen lisäksi Suomessa on joillakin yliopistoilla käytössä Elsevierin PURE, joillakin taas kotimainen SoleCRIS.

Kaupallisten sovellusten ohella italialaisen Cinecan kehittämä DSpace-CRIS-moduli voi olla jatkossa potentiaalinen vaihtoehto etenkin pienemmille organisaatioille. Moduli on tosin yhä melko varhaisessa kehitysvaiheessa, eikä Kansalliskirjasto ole ainakaan vielä ennättänyt selvittää, miten suuria haasteita sen käyttöönottoon liittyisi.

3.7 Tilastointi

AMKIT- konsortiolla on tarvetta organisaatiokohtaisille selkeille tilastoille aineistojen määristä ja käytöstä. Todettiin, että organisaatiokohtaiset lataus- ja kartuntamäärät ovat jo nyt saatavissa Theseuksesta. Kartuntamäärien osalta kyseessä ei ole varsinaisesti tilasto, mutta tiedot pystyy hakemaan järjestelmästä. Range-hakua on parannettu viime aikoina. Lisäksi on mahdollista saada tietoa tiedostojen viemästä levytilasta yms.

Kartuntatietojen saamiseksi voisi olla hyödyllistä tehdä selkeämpää ohjeistusta ja mahdollisesti työkaluja. Kansalliskirjastossa on kehitetty visualisointityökalua tilastoille, mutta se ei tuo lisätietoja organisaatiotasolle. Google Analytics -sovelluksessa organisaatiokohtaisuutta on vaikea toteuttaa ilman omaa ulkoasuteemaa. Tietoja saa tällöin vain arkistokohtaisesti. DSpace 5:ssä tarjoaa välineitä parempaan Google Analytics -kytkentään. Se mahdollistaa ilmeisesti myös tietojen keräämisen pdf-latauksista.

4. OpenAire -yhteensopivuus

OpenAire-yhteensopivuus ei ole enää DSpace 3:n käyttöönoton jälkeen teknisesti hankala asia. Se vaatii kuitenkin myös yhtenäistä metadatta, jossa on tarvittavilta osin OpenAiren haluamat spesiaalimuotoiset kentät. Lisäksi tarvittaisiin poliittisia linjauksia sen suhteen, riittääkö esim. Doriassa vain yksi OpenAire-setti vai tarvitaanko erillisiä organisaatiokohtaisia settejä? Mitä aineistoja halutaan mukaan? Epäselvää on myös mistä julkaisuista OpenAire oikeastaan haluaa kerätä tietoja. OpenAiren fokus on laajentunut niin, että se ei ole enää tarkoitettu vain EU-rahoitteisissa projekteissa tuotetuille julkaisuille.

Asia on kansallisesti vielä kesken myös muuten kuin KK:n ylläpitämien palveluiden osalta. OpenAire -asioista pitäisi keskustella yleisellä tasolla sekä tässä ryhmässä että muualla. Haasteena on se, että on julkaisuarkistoinstansseja, joissa on usean organisaation aineistoja. Keskustelua pitäisi käydä erityisesti sen suhteen miten sisältö saadaan organisoitua sekä metadata- että settitasolla. Riittääkö, että julkaisut löytyvät OpenAiresta vai pitääkö niiden olla löydettävissä organisaatiokohtaisesti? Halutaanko OpenAireen myös opinnäytteet vai vain artikkelit? OpenAire-yhteistyötä koordinoi kansallisella tasolla Helsingin yliopiston kirjasto.

5. Muut Asiat

Keskusteltiin seuraavan kokouksen ajasta. Seuraava kokous on syksyllä. Viikolla 34 perjantaina 21.8. sekä aamupäivä ja iltapäivä näyttäisivät tällä hetkellä mahdollisilta vaihtoehdoilta, mutta täytyy varmistaa sopiiko päivä myös ryhmän muille jäsenille. Mikäli ajan kanssa on hankaluuksia, asiasta lähetetään Doodle-kysely osanottajille myöhemmin.