

2018-09-25

Julkaisuarkistojen asiantuntijaryhmä, 25.09.2018

Julkaisuarkistojen asiantuntijaryhmä, 25.09.2018

1. Kokouksen avaus
2. Uudet palvelut ja muu kehitystyö
3. Julkaisuarkistot ja GDPR
4. Kesän konferenssit
5. Muut asiat
6. Seuraava kokous

Aika: 25.09.2018 klo 10.00-12.00

Paikka: Fabianian neuvotteluhuone, 2. krs, Yliopistonkatu 1 ja AC-etyhyteys

Läsnä: Jyrki Ilva (Kansalliskirjasto), Lassi Lager (Kansalliskirjasto), Minna Liikala (Terveysten ja hyvinvoinnin laitos), Minna Marjamaa (AMKIT-konsortio), Mikko Pennanen (Turun yliopisto), Päivi Rosenström (Kansalliskirjasto), Tanja Vienonen (Kansalliskirjasto), Samu Viita (Kansalliskirjasto)

Etänä: Ilkka Nokelainen (Lappeenrannan teknillinen yliopisto), Joonas Parviainen (Maanpuolustuskorkeakoulu), Silja Pirkola (Luonnonvarakeskus), Toni Raja-Hanhela (Lapin yliopisto), Iris Tahvanainen (Lappeenrannan teknillinen yliopisto), Tiina Tolonen (AMKIT-konsortio)

1. Kokouksen avaus

- Puheenjohtaja avasi kokouksen klo 10.07

2. Uudet palvelut ja muu kehitystyö

- Lappeenrannan teknillisen yliopiston uusi [LutPub-julkaisuarkisto](#) avattiin kesäkuussa
 - LutPubin takautuvat sisällöt siirretty Doriasta
 - Löytyvät vielä jonkin aikaa kahdesta paikasta, mutta poisto Doriasta on tarkoitus sopia Kansalliskirjaston kanssa piakkoin
 - Poiston jälkeen LutPubiin siirretään myös takautuvat tilastot
 - Converis-kytkentää koskeva selvitystyö on vielä tekemättä - tarvitaan lisätietoja Converiksen ominaisuuksista
 - Miika Nurmisen (Jyväskylän yliopisto) [CRIS2018-konferenssiesityksen](#) perusteella kytkennässä voi olla haasteita
 - Selvitystyö aloitettiin keväällä, on edelleen samassa tilassa
 - Yksi avoin kysymys on millä tekniikalla siirto onnistuu järjestelmästä toiseen
 - Rajapintana voisi käyttää integraatioalustaa, kuten [Windowsin BizTalkia](#)
 - Converiksen DSpace-kytkentä perustuu ilmeisesti SWORDiin, mutta sen käytössä on ollut Jyväskylässä ongelmia
 - Siirrettävästä metadatasta päätetään vasta sen jälkeen, kun rajapinnan toiminnasta saadaan selvyys
 - Kytkentä olisi erityisen tarpeellinen artikkelien rinnakkaistallennuksessa, joka tällä hetkellä tehdään käsityönä ja vie jo valmiiksi vähäisiä resursseja
 - Turun yliopistolla voi olla kiinnostusta Converis-kytkentään, jos sen kehitys etenee
- Theseuksen tulevat uudistukset
 - Syöttölomakkeiden uusiminen on suunnitteilla, tarkoitus tehdä samaan aikaan kokoelmarakenteen uudistuksen ja metadatan päivityksen kanssa
 - Tiina Tolonen on katsonut metadattaa, Minna Marjamaa on aloittamassa työn pian
 - Laureassa tehdään paljon rinnakkaistallennuksia ja siellä toivotaan, että syöttölomake olisi mahdollisimman yksinkertainen ja helposti ymmärrettävä tietojen syöttäjälle, joka on usein muu kuin julkaisuarkistoammattilainen
 - Toivottavaa olisi, että metadatan tarkastaja voisi tehdä tarvittavat korjaukset ja lisäykset myöhemmin
 - Justus-integraation testausta ei ole vielä aloitettu, CSC odottaa Theseuksen metadattaa koskevia ratkaisuja
 - Justuksen lomakkeen käytössä on ollut toistaiseksi vaikeuksia
 - Ongelmia on ollut mm. Eduuniin kirjautumisessa
 - Metadatan syötön tulisi olla helppoa; julkaisujen syöttäjät ovat olleet puutteista yhteydessä CSC:hen
 - Workflow vaatii vielä kehitystyötä
 - Theseuksen uusi ulkoasu on saatu graafikolta, lähetetty Samu Viidalle ja Anis Moubarikille Kansalliskirjastossa
 - Aiemmin on sovittu, että jos graafinen materiaali on valmis ennen syyskuun loppua, uusi ulkoasu saadaan julkaistua marraskuussa
 - Asiaa käsitellään myös AMK-kirjastojen johtajien kokouksessa
 - Theseuksen yhteydessä keskusteltiin myös pitkäaikaissäilytykseen liittyvistä kysymyksistä
 - Aija Vahtolalla on saatu tieto, ettei asia ole edennyt

- PDF/A-konversion osalta ollaan odoteltu PAS-asian etenemistä: konvertterin rakentaminen ei ole hyödyllistä ennen kuin tiedetään onko se tarpeen
 - Theseuksessa on monenlaisia tallentajia, joten ohjeistuksen pitäisi olla niin yksiselitteinen, että kaikki osaavat toimia samalla tavalla
- Kokoelmarakenteen yksinkertaistaminen on suunnitteilla myös Jukurissa
 - Työ on aloitettu, Samu Viita ja Tanja Vienonen Kansalliskirjastosta menevät vierailulle LUKE:n Viikin pisteeseen lokakuussa
 - Haasteita on, mutta kokoelmarakenteen yksinkertaistaminen on hyvä tavoite julkaisuarkiston ylläpidonkin kannalta
 - Kun mahdollisimman iso osa julkaisujen tiedoista on metadatasissa kokoelmarakenteen sijaan, tulevat päivitykset on helpompi tehdä
 - Kokoelmia pitää olla aina ainakin yksi, kustomoidut syöttölomakkeet ja käyttörajat aineistot lisäävät kokoelmien määrää
 - Takautuvan datan kanssa voi olla haasteita
 - Dataa on tuotu eri organisaatioista, vanhasta Trip-tietokannasta ja aiempien tutkimuslaitosten (esim. Metla) tietokannoista
 - Jukuri toimii sekä julkaisuarkistona että julkaisurekisterinä
 - Tanja Vienonen tutustuu Jukurin metadatasien ennen kokousta
 - Turun yliopistossa on tehty vastaava uudistus eikä negatiivista palautetta ole saatu; jos metadata on laadukasta, kokoelmarakenteen ei tarvitse olla monitahoinen
 - Muutoksen yhteydessä on hyvä pyytää organisaation henkilökuntaa tarkistamaan verkkosivuille ja muualle linkitettyjen julkaisujen osoitteiden toimivuus ja muistuttaa **pysyvien tunnisteiden** käytöstä
- Julkaisuarkistojen ja Justuksen kytkentä
 - Odotellaan Justuksen uutta testiversiota, jota on lupailtu CSC:ltä myöhemmin tänä syksynä
 - AMK:ien osalta puutteiden ja toivottujen ominaisuuksien esille nostaminen olisi hyvä tehdä nyt, jotta ne saadaan käyttöön Theseuksen uudistuksen yhteydessä
 - Justuksen käyttöönotto ei toteutune Theseuksen osalta tänä syksynä, sillä ehdotettu ajankohta (joulukuu) ei ole ihanteellinen opinnäytteiden syöttökiireen vuoksi
- DSpace-syöttölomakkeiden uudet ominaisuudet
 - Kytkenä Finto-rajapintaan ja URN-generaattoriin on toteutettu; enää ei tarvitse hyppiä eri sivuilla hakemassa näitä tietoja
 - Syöttölomakkeiden määrittelymahdollisuus taulukon avulla
 - DSpaceen voidaan hakea lomakkeiden metadatat Google Driveen luotavasta taulukosta, josta tarvittavat tiedot pyöräytetään skriptin avulla sisään julkaisuarkistojärjestelmään
 - Käytössä mm. Julkariissa
 - Taulukoiden käytöstä on suuri apu julkaisuarkiston ylläpitäjälle, kun pyydyt muutokset eivät huku sähköpostin syövereihin
 - Pudotusvalikoiden päivittämismahdollisuus toisella palvelimella olevien listojen avulla
 - Myös Theseuksessa voidaan todennäköisesti siirtää käyttämään DSpace-syöttölomakkeita tulevan uudistuksen yhteydessä
 - Lomake ei olisi aivan samanlainen kuin nykyiset lomakkeet, mutta esim. kieliversiointi onnistuu (esim. opinnäytteen taso -kenttä)
 - Shibboleth-tunnistautumista testataan tiedon syöttäjien kirjautumisessa
 - Lapin yliopiston Lauda toimii ensimmäisenä pilottina, muissa voidaan ottaa käyttöön pilotin jälkeen
- Julkaisuarkistoon tallennettujen kuvien esittämiseen uusi galleria-toiminto
 - Katso [esimerkki Doriasta](#)
 - Toimii tietueissa, joihin on tallennettu useampi kuva, joita voi galleriassa selailla
 - Toimintoa jatkokehitetään vielä
 - Mobiilikäyttö optimoidaan
 - Pohdinnassa näytetäänkö esikatselukuvat tietueen etusivulla kaikista
 - Kuvien esitysjärjestys riippuu tiedostojen järjestyksestä; DSpace:ssa voi järjestää kuvat uudelleen tiedostojen järjesty -osiossa
- Metadatasuunnittelu
 - Tulossa yhteistapaaminen joissain julkaisuarkistoissa käytössä olevan automaattisesti generoituvan viittauslaatikon tietojen määrittelystä
 - Doodleen saatu jonkin verran vastauksia
 - Tärkeää tässä on yhtenäisen metadatan tallennusformaatin käyttö
 - Tapaaminen järjestetään torstaina 11.10. klo 10-12, kutsu lähetetään ilmoittautuneille
 - Toivotaan mahdollisimman paljon osallistujia paikan päälle, mutta osallistuminen onnistuu etänäkin
 - Tapaamisessa on tarkoitus ainakin sopia minimivaatimuksista, joilla laatikko muodostetaan
 - Minimivaatimukset pitää sopia niin metadatatenttien määrän kuin niiden sisällön osalta (esim. montako merkkiä arvossa pitää olla)
 - Lisäksi on mietittävä halutaanko, että viittauslaatikko noudattaa jotain tiettyä formaattia/standardia ja mitä tietoja tätä varten tarvitaan

3. Julkaisuarkistot ja GDPR

- Asiaan liittyvä selvitystyö on jatkunut Kansalliskirjaston eri palveluissa
 - Julkaisuarkistot ovat yksi palanen muiden palveluiden joukossa

- Tulkinat eri osapuolten vastuista ovat tarkentuneet prosessin aikana
- Heli Kautonen esitteli [sopimusliiteluonnosta](#) tiedonhallinnan ohjausryhmässä 11.9.18
 - Käsiteltiin luonnokseen tehtyjä muutoksia (merkitty liitteeseen punaisella tähdellä)
 - Palveluiden henkilötiedot (jaetuissa julkaisuarkistoissa), sekä palveluiden käyttäjistä kerätyt lokitiedot ovat Kansalliskirjaston vastuulla
 - Asiakasorganisaatiot vastaavat omasta henkilötietorekisteristä, sekä metatiedoista ja aineistossa olevista henkilötiedoista
- Luonnos sopimusten tietosuojaliitteistä on lähtenyt kommenttikierrokselle
 - Jaana Routakangas Kansalliskirjaston Tietovarannot-yksiköstä on lähettänyt luonnoksen julkaisuarkistojen sopimussyhteyshenkilöille
 - Kommentointiaikaa on 15.10.2018 asti
 - Julkaisuarkistoissa haasteena ovat jaetut arkistoinstanssit
- Theseuksen tietosuojailmoitukset on lähetetty käännettäväksi; voi olla, että ilmoitukseen tulee muutoksia vuodenvaihteen jälkeen
- Kansallinen lainsäädäntö tietosuojasta on vielä hyväksymättä

4. Kesän konferenssit

- [Open Repositories 2018](#) Bozemanissa, Coloradossa
 - Jyrki Ilva esitteli konferenssimatkaa, katso [esitys](#)
 - Tämän vuoden teemana oli "Sustaining Open", joka näkyi myös esityksissä – aiheina mm.
 - Kaupalliset vs. yhteisön voimin avoimesti tuotetut palvelut ja ohjelmistot
 - Keskitetyt vs. paikalliset palvelut
 - Palvelujen pitkän tähtäimen ylläpito
 - Konferenssiin osallistui n. 350 henkeä, joka on melko tavallinen määrä, kun konferenssi järjestetään Pohjois-Amerikassa
 - Tim Donohue esitteli DSpace 7:n kuulumisia
 - Näillä näkymin uusi versio julkaistaan alkuvuodesta 2019, vaikka monet asiat ovat vielä kesken
 - Version uutuuksia mm. uusi käyttöliittymä, kokonaan uusittu REST API, ResourceSync-rajapinta ja tuki koneluettavan metadatan välittämiseksi http-headerissa ("signposting")
 - Tietomalliin tehdään muutoksia, joiden toteuttaminen on viivästyttänyt kehitystyötä
 - Tulevaisuudessa voidaan julkaisujen lisäksi kuvailla muitakin entiteettejä, mm. tekijöitä tai esim. lehtimuotoisia julkaisuja
 - Kehitetään eteenpäin tulevissa DSpace-versioissa
 - Tietomallissa on tehty DSpace-CRISin tietomallia vastaavia ratkaisuja, mutta tekninen toteutus on erilainen
 - Esityksen mielenkiintoisinta antia oli nähdä miten syöttölomaketta on kehitetty
 - Kokouksessa kommentoitiin nykyisen REST-rajapinnan käytön haastavuutta, sillä laajasti käytettynä sen kanssa on ilmennyt ongelmia
 - DSpace 7:n uusia ominaisuuksia pohdittiin käytön kannalta
 - ennen käyttöönottoa pitää varmistaa sekä toimivuus, että oikeudelliset asiat, kuten tietosuojakysymykset
 - Olisi hienoa, jos julkaisuarkistoihinkin voisi kytkeä auktoriteettitietokannan Melinda-metatietovarannon tapaan
 - Pitää myös varmistaa kuinka uudistukset kytkeytyvät toisiinsa ja muihin versioihin
 - Ensi vuoden OR-konferenssi järjestetään 10.-13.6.2019 Hampurissa
 - Toivotaan mahdollisimman laajaa osallistumista suomalaisista julkaisuarkistoista
 - Call for Proposals avataan lokakuun lopulla
- [CRIS2018](#) Uumajassa
 - Osallistujissa enemmän suomalaisia kuin OR:ssä, yhteensä toistakymmentä, kokonaisuudessaan konferenssiin osallistui n. 150 henkeä
 - DSpace-CRIS oli näkyvästi esillä
 - Jopa enemmän esillä kuin kaupallisten toimijoiden ohjelmistot
 - Kehitetään italialaisessa 4Science-yrityksessä, aiemmin Cinecassa ("Italian CSC")
 - Kysymyksiä herättää se, että DSpace-CRIS on yksi DSpacen kehityshaara, ei erillinen moduuli
 - Suomessa DSpace-CRISiä ei ole toistaiseksi otettu käyttöön
 - Haasteena ollut eri käyttöliittymä (JSPUI) kuin Suomessa käytetyissä DSpace-versioissa (XMLUI)
 - Tämä ongelma poistuu DSpace 7:n uuden käyttöliittymän myötä
 - Todettiin, että CRIS-järjestelmä tarjoaisi organisaatioille uusia mahdollisuuksia
 - Voisi lisätä kiinnostusta myös OA:ta kohtaan
 - DSpace 7:ään on tulossa myös mahdollisuus toteuttaa tutkijoiden omia sivuja

5. Muut asiat

- Julkaisuarkistojen sisältämien kirja-aineistojen metatietojen hyödyntäminen Melindassa, katso [esitys](#)
 - Lassi Lager esitteli suunnitelmia hyödyntää metatietovaranto Melindassa valmiita kuvailutietoja muista tietokannoista
 - Melindaan ei ole tarkoitus haravoida kaiken aineiston metatietoja, vaan vain ISBN-tunnuksellisia aineistoja (esim. väitöskirjat)
 - Vuosittain Melindaan voitaisiin saada noin 3000 uutta kirjojen kuvailutietuetta

- ISBN-tunnukselliset julkaisut ovat useimmiten julkaisuarkistoasiantuntijoiden tallentamia, joten metatiedot ovat todennäköisesti oikein
 - Jatkossa haravointi on tarkoitettu saadaan automatisoitua säännöllisesti tehtäväksi prosessiksi
 - Alussa ihmistarkastus välissä
 - Metatiedot haravoidaan suoraan Melindaan, josta ne replikoiduvat Fennica-kansalliskirjastojaan, tarvittaessa muihin (kirjasto)tietokantoihin ja Fennicaa /kirjastotietokannoista Finnaan
 - Finnan tuplakontrollissa voisi käyttää URN-tunnuksia hyödyksi
 - Aluksi haravoidaan vain metatieto, mutta myöhemmin haravointiin liitetään myös kokotekstitiedostot vapaakappaleprosessin kautta
 - Ehdottomasti hyvä asia, koska väitöskirjojen haku voi olla hankalaa, kun ei ole yhtä luotettavaa paikkaa mistä niitä voisi etsiä
 - Mainittiin, että väitöskirjat ilmoitetaan OKM:n vuosittaisessa julkaisutiedonkeruussa, joten ne ovat Juulissa melko kattavasti (aivan uusimmat haettava eri tietokannoista)
- Muutoksia Kansalliskirjastossa
 - Vuoden 2019 alusta toteutuva organisaatiouudistus vaikuttaa myös julkaisuarkistopalveluihin
 - Edellinen uudistus oli 2017
 - Nyt suuret yksiköt jaetaan pienemmiksi palvelujen mukaan
 - Julkaisuarkistopalvelut siirtyvät Tietovarannoista Julkaisemisen palvelut -yksiköksi, jonka esimieheksi siirtyy Heli Kautonen entisestä Asiakkuuden hallinta -yksiköstä
 - Nina Hyvönen jatkaa toisen yksikön esimiehenä
 - Leipätehtaan tiloista luovutaan tammikuussa 2019
 - Myös Kansalliskirjaston ylikirjastonhoitaja on vaihtumassa edellisen johtajan jäätyä eläkkeelle
- Julkaisuarkistopalveluiden asiantuntijaryhmän tulevaisuus
 - Nykyisen ryhmän kausi päättyy, uuden ryhmän kokoonpanosta päätetään syksyn aikana
 - On mahdollista, että osa ryhmästä joudutaan kutsumaan erikseen
 - Edellisellä kerralla SYN (nyt FUN) ei katsonut tehtäväkseen nimetä yliopistokirjastojen edustajia, koska palvelun asiakkaina on vain osa yliopistokirjastoista
 - Kaikkia ryhmän jäseniä ei ole pakko vaihtaa kerralla
 - Ryhmän toimintakausi voisi olla kaksi vuotta, jotta kaikki pääsevät osallistumaan
 - Iris Tahvanainen: On hyvä, että ryhmä jatkaa, sillä se on ollut tarpeellinen ja julkaisuarkiston kehittämistyössä erittäin hyvä tietolähde
 - Myös Kansalliskirjaston asiantuntijoiden mielestä ryhmä on toiminut hyvin tiedon välittämisessä asiakasorganisaatioille
 - Yleisemmälle, kansalliselle ryhmälle olisi voinut myös olla tarvetta
 - Ryhmän toimintaan voisivat osallistua kaikki Kansalliskirjaston julkaisuarkistoasiakkaat
- Avoin Tiede -koordinoiti on siirtynyt Tieteellisten Seurain Valtuuskunnalle (TSV)
 - Kansalliskirjastossa ja ehkä myös muissa organisaatioissa tarvittaisiin tähän enemmän resursseja
 - Yliopistokirjastoilla on ollut epävirallinen yhteistyöryhmä OA-verkosto
- Tiedottamista ja yhteistyötä julkaisuarkistoasioissa voisi auttaa vuosittain järjestettävä seminaari tai vastaava
 - Organisaatiouudistuksen jälkeen voisi selvittää löytyisikö tällaisen järjestämiseen resursseja Kansalliskirjastosta
 - Kommentoitiin, että olisi kiinnostavaa kuulla muidenkin kuin Kansalliskirjaston asiakkaiden kehittämistä uudistuksista ja niiden onnistumisesta
 - Esim. Jyväskylän yliopiston Converis-kytkennästä
- Theseuksen kymmenvuotisseminaari järjestetään 21.11.18 Tampereella
 - Ohjelmaa suunnitellaan, tulee saataville myöhemmin
 - Kansalliskirjastosta pyydetty esityksiä Jyrki Ilvalta ja Samu Viidalta
 - CSC:ltä esityksen pitää Justus-kehityksestä vastaava Joonas Nikkanen
 - Myös johtajatasen esitys on suunnitteilla
 - Ei etäyhteyttä, sillä keskiössä on eri AMK:ien työntekijöiden tapaaminen kasvokkain

6. Seuraava kokous

- Vuoden viimeinen kokous järjestetään **tiistaina 6.11.2018 klo 10-12** Sörnäisissä Leipätehtaalla kirjastoverkkopalvelujen tiloissa ja etäyhteydellä