

# 2016-12-15

## Julkaisuarkistopalveluiden asiantuntijaryhmä, 15.12.2016

Julkaisuarkistopalveluiden asiantuntijaryhmä, 15.12.2016

1. Metadatan yhdenmukaistaminen
  2. DSpace-versiopäivitykset
  3. Ajankohtaista palveluista
  4. Kuulumisia DSpace-ohjausryhmästä
  5. Muut asiat
- Seuraava kokous: 17.3. klo 10-12 Kirjstoverkkopalveluissa ja etäyhteydellä.

**Aika:** 15.12.2016 klo 12.00-14.00

**Paikka:** Leipätehdas, Kardemumma V503, Kaikukatu 4, 5. krs ja AC-etäyhteys

**Puheenjohtaja:** Mikko Pennanen

**Sihteeri:** Tanja Vienonen

**Läsnä:** Nina Hyvönen (Kansalliskirjasto), Jyrki Ilva (Kansalliskirjasto), Lassi Lager (Kansalliskirjasto), Mikko Pennanen (Turun yliopisto), Päivi Rosenström (Kansalliskirjasto), Iris Tahvanainen (Lappeenrannan teknillinen yliopisto), Tanja Vienonen (Kansalliskirjasto)

**Etänä:** Minna Marjamaa (AMKIT-konsortio), Ilkka Nokelainen (Lappeenrannan teknillinen yliopisto), Joonas Parviainen (Maanpuolustuskorkeakoulu), Toni Raja-Hanhela (Lapin yliopisto), Mikko Saunamäki (Luonnonvarakeskus), Tiina Tolonen (AMKIT-konsortio)

**Poissa:** Minna Liikala (THL), Teemu Makkonen (AMKIT-konsortio), Samu Viita (Kansalliskirjasto)

### 1. Metadatan yhdenmukaistaminen

- [TAJUA-yhteistyöverkoston wiki-sivu](#)
- Yhdenmukaistamistyö jatkuu vuodenvaihteen yli, todennäköisesti suositus on valmis alkuvuodesta 2017.
- Marraskuussa Helsingissä järjestetty tapaaminen muiden organisaatioiden metadata-asiantuntijoiden kanssa oli onnistunut ja saatu palaute oli positiivista. Tällaisia tapaamisia toivottiin lisää.
- Mikko Saunamäki: Metadatasuosituksen kokoaminen on tervetullutta kehitystä.
  - Kysymys käytännön toimista: koordinoiko Kansalliskirjasto metadatan yhtenäistämistä vai miten asia etenee?
  - Suosituksen jalkauttaminen on yksi Tajua-projektin tehtävistä, Tanja Vienonen lähtee vierailemaan organisaatioissa tai esittelemään suositusta kirjastojen tapahtumiin.
  - Pääasiassa vastuu metadatan siivoamisesta on ylläpitäjillä, joita Kansalliskirjasto neuvoo ja auttaa tarvittaessa.
- Iris Tahvanainen: Lappeenrannasta tapaamisessa olleet Kaisa Eloranta ja Ilkka Nokelainen olivat iloisia että yhtenäistämistä tehdään Kansalliskirjastossa, sillä tietotaitoa löytyy eniten sieltä.
- Metadatan laatu vaihtelee julkaisuarkistoittain, minkä lisäksi on mietittävä miten aggregointi Finnaan ym. hoidetaan.
- Asiaa käsitellään vielä tulevilla kokouksilla.

### 2. DSpace-versiopäivitykset

- Kaikki Kansalliskirjaston ylläpitämät julkaisuarkistoinstanssit päivitetään DSpacen versioon 5.6.
- Kiirettä lisää se, että joissain nykyisissä instansseissa käytössä olevaan CentOS5-käyttöjärjestelmään ei enää tule päivityksiä - tuki loppuu maaliskuussa 2017.
- Päivityssuunnitelmaa lähdetään toteuttamaan alkuvuodesta 2017.
- Alpha-palvelin (julkaisuarkistoalusta) on jo olemassa mutta vaatii vielä työtä tätä monistetaan, minkä jälkeen se toimii kaikkien julkaisuarkistojen palvelimena.
- Luonnonvarakeskuksen Jukuri-julkaisuarkisto päivitetään ensimmäiseksi juuri CentOS5-palvelimen vaihtamisen vuoksi, muu aikataulu on vielä sopimatta.
  - Erityisesti Doria ja Theseus vaativat enemmän valmistelua moniulotteisuutensa vuoksi.
- Mikko Saunamäki: Jukuri toimii myös julkaisurekisterinä, joten versionvaihto ei saa häiritä alkuvuodesta tehtävää MMM:n ja OKM:n julkaisutiedonkeruuta.
  - MMM-tiedonkeruun päivämäärä vaihtelee, mutta määräpäivä on yleensä helmikuun puolivälin paikkeilla.
  - OKM-tiedonkeruu on suurin piirtein samaan aikaan.
- Julkaisuarkistoa voi käyttää koko päivitysprosessin ajan, mahdolliset katkot tallennuksessa kestävät muutamia päiviä, eivät missään nimessä viikkoja.
  - Katkoja tulee vain tallennukseen, päivitys ei näy muussa käytössä.

- Päivitys koordinoidaan yhdessä asiakkaiden kanssa, apua tarvitaan erityisesti jos julkaisuarkistossa on oma domain käytössä.
- Iris Tahvanainen: Myös opiskelijoiden valmistumisprosessi on otettava huomioon, ettei tallennuskatko osu kiireisimpään aikaan.
  - Kirjaston pitää tarkistaa opinnäytetyön kokotekstin löytyminen julkaisuarkistosta viimeistään kolme päivää sen syöttämisen jälkeen.
- Ennen päivitysten aloittamista on vielä selvítettävä mitä ominaisuuksia DSpace5:sta otetaan käyttöön.
  - Mobiilioptimoitu Mirage 2 otetaan todennäköisesti käyttöön vasta vähän myöhemmin, sillä se vaatii mm. julkaisuarkistoissa käytettyjen teemojen muokkaamista.
- Uudempi DSpace-versio ei vaikuta handle- tai muihin tunnisteisiin eikä OAI-PMH -rajapintamuunnoksiin.
  - OAI-PMH:n kautta välitettävän datan formaatti testataan vielä ennen käyttöönottoa.
- Tutkittavaa on vielä ennen kuin Alpha-palvelinta voidaan ottaa käyttöön.
  - Apache Lucene ja Solr vs. OpenSearch -haku on vielä selvittämättä.
- Tiedotus päivityksestä alkaa ensi vuoden alussa, julkaisuarkistot päivitetään yksi kerrallaan.

### 3. Ajankohtaista palveluista

Theseuksen kaatumisongelma ratkaistu

- Theseus oli syksyllä pahoissa vaikeuksissa, pahimmillaan palvelimet kaatuivat viiden minuutin välein.
- Kyse oli useasta päällekkäisestä ongelmasta: toisaalta monet palvelimia kuormittavat haravointiohjelmat olivat aktivoituneet samanaikaisesti, minkä lisäksi virheilmoitussivuihin liittyvät tekniset ongelmat lisäsivät kuormaa merkittävästi.
- Tilanne saatiin lopulta hallintaan: palvelinten optimointi ja virheilmoitusten käsittelyn parantaminen auttoivat.
- Asiasta on ilmoitettu DSpace-kehittäjille, mutta ongelma ei välttämättä poistu versiopäivityksellä, sillä vika on syvemmällä.
- Ongelmasta oli paljon häittää ammattikorkeissa, kirjastot saivat paljon palautetta toisaalta Theseus on otettu käyttöön jo vuonna 2009 ja tämä oli ensimmäinen vika.
- Theseus on Kansalliskirjaston ylläpitämistä julkaisuarkistoista suurin niin kooltaan kuin käyttäjämäärältään, minkä vuoksi ongelmat näkyvät siinä ensimmäisenä.
- Tilastojen poistaminen käytöstä oli yksi yritys löytää ongelmien perussy, siksi ne olivat poissa käytöstä muutamia päiviä.

Julkaisuarkistopalveluiden henkilöresurssit lisääntyvät

- Tajua-hankkeessa tällä hetkellä työskentelevä tietojärjestelmäasiantuntija Anis Moubarik on valittu uudeksi julkaisuarkistopalvelujen vakituiseksi työntekijäksi.
- Hän on ollut töissä Kansalliskirjastossa noin kaksi vuotta ja oli mukana kehittämässä esim. SimpleREST-rajapintaa.
- Kokouksen osallistujat olivat yhtä mieltä siitä, että resurssien lisääntyminen on hyvä asia.

### 4. Kuulumisia DSpace-ohjausryhmästä

Jyrki Ilva on ollut syyskuusta lähtien mukana DSpacen Steering Groupissa, joka koostuu DSpacen tukijaorganisaatioiden edustajista.

- Kansalliskirjasto on tukenut DuraSpacen toimintaa taloudellisesti jo muutaman vuoden ajan.
- Ryhmän puheenjohtajana toimii David Lewis Indiana University Purdue University Indianapolis (IUPUI) University Librarysta.

Syksyn aikana puheenaiheina ryhmässä olivat etenkin DSpace 7:n kehitystyö ja DSpacen rahoitustilanne.

- Fedora-ohjelmistolla on kaksinkertaiset tulot DSpaceen verrattuna, sillä Fedoraan panostetaan erityisesti Yhdysvaltojen isoissa yliopistoissa, jotka ovat merkittävimpiä rahoittajia.
- DSpace-pohjaisia julkaisuarkistoja on paljon Euroopassa, Aasiassa, Etelä-Amerikassa ja Afrikassa, mutta tukijoita on vähemmän kuin itse käyttäjiä Euroopassa vain 10.
- Tukijoita tarvittaisiin enemmän, sillä myös avoimen lähdekoodin kehitystyöhön tarvitaan resursseja.
  - Kehitystyötä voi tukea niin taloudellisesti kuin lahjoittamalla työtunteja kehittämistyöhön.
  - Kansalliskirjasto on nostanut oman tukisummansa ainakin toistaiseksi 10 000 dollariin ajatuksena on tukea DSpacea myös asiakasorganisaatioiden puolesta.

Avoimen lähdekoodin ohjelmien kehitystyö liittyy avoimeen tieteseen ja se on mainittu ATT-hankkeen dokumenteissakin, sillä myös käytettyjen työkalujen tulisi olla avoimia.

DSpaceCRIS-moduuli on yksi DSpacen potentiaalisista tulevaisuuden vahvuuksista: kehittäminen tarvitsisi tarpeeksi laajan tukijoukon taakseen, jotta työ ei jää kesken.

- On kiinnostavaa nähdä mitä tapahtuu Italiassa DSpaceCRISin kehitystyö on aktiivista.
- Toistaiseksi haasteena on ollut se, että CRIS toimii vain DSpacen JSP-käyttöliittymän kanssa eikä ole yhteensopiva Suomessa käytetyn XML-käyttöliittymän kanssa.
- Tältä osin tilanne muuttuu DSpace 7:n ja uuden käyttöliittymän myötä.
- Lisäksi myös DSpacen tietomalli kaipaa kehittämistä, jotta CRIS ei irtaudu DSpacen kehityslinjasta omaksi erilliseksi haarakseen - mahdollisesti DSpace 7:ssä.
- DSpaceCRIS on käytössä mm. euroCRIS'in julkaisuarkistossa: <http://dspacecris.eurocris.org/>.

## 5. Muut asiat

Kansalliskirjasto järjestää 25.1.2017 rinnakkaistallennusta käsittelevän iltapäiväseminaarin yhteistyössä Jyväskylän yliopiston kirjaston kanssa.

- Tapahtumaa voi seurata etäyhteydellä tai paikan päällä Fabianian auditoriossa.
- Ohjelma: <https://www.kansalliskirjasto.fi/fi/tapahtumat/rinnakkaistallentaminen-suomessa-iltapaivaseminaari>
- Ilmoittautuminen: <https://elomake.helsinki.fi/lomakkeet/76184/lomake.html>

Iris Tahvanainen: Arkistolaitoksen päätös opinnäytteiden pitkäaikaissäilytyksestä on tervetullut, mutta toiminta käytännössä on mietittyänyt.

- Esa-Pekka Keskitalo on laatimassa FAQ-sivua aiheeseen liittyen, sillä kysymyksiä on tullut monesta organisaatiosta.
- Kansalliskirjasto varmistaa vielä Arkistolaitoksen kannan.
- Ammattikorkeissa ei olla vielä luovuttu elektronisten opinnäytteiden paperiversioiden säilyttämisestä, sillä on epävarmuutta siitä voiko paperiarkiston kokoamisen todella lopettaa.
- Kansallisarkistolta on vaikeaa saada suoraa vastausta asiaan.
- Kansalliskirjasto pyrkii selvittämään asian niin, että Kansallisarkistolle lähetetään ehdotus opinnäytteiden pitkäaikaissäilytyksestä ja kysytään hyväksyvätkö he ehdotuksen.

## Seuraava kokous: 17.3. klo 10-12 Kirjastoverkkopalveluissa ja etäyhteydellä.

Aiheita: Julkaisuarkistosopimusasiat, metadata ja DSpacen päivityksen eteneminen.