

2019-10-02

- 1. Kokouksen avaaminen ja asialistan hyväksyminen
- 2. Tiedotusasioita
- 3. Edellisen kokouksen muistio
- 4. Theseuksen käyttökatkot syksyllä
- 5. PAS-tilannekatsaus
- 6. Sisältöjen nostot etusivulle
- 7. Justuksen kehittäminen
- 8. Theseuksen ylläpito ja kehittäminen
- 9. Kustannukset vuonna 2020
- 10. Muita mahdollisia asioita
- 11. Seuraava laatupalaveri
- 12. Kokouksen päättäminen

Aika: 2.10.2019 klo 13–15

Paikka: Skype

Osallistujat:

Hanna Lahtinen, puheenjohtaja
Minna Koistinen
Tiina Tolonen
Antti Nyqvist
Minna Marjamaa
Jyrki Ilva
Samu Viita
Päivi Rosenström
Reeta Kuuskoski, sihteeri

1. Kokouksen avaaminen ja asialistan hyväksyminen

Esiteltiin uusi kansalliskirjastolainen Reeta Kuuskoski. Reeta valittiin sihteeriksi.

2. Tiedotusasioita

Arene on solminut Teoston kanssa Mediapienlupan Theseukseen tallennetun musiikin käyttöä varten. Mediapienlupa on lisäksi liitetty Piensynkronointilupa, joka antaa luvan käyttää AV-tuotantoihin liitettyä musiikkia. Syöttölomakkeelle on määritetty tätä varten uusi kenttä, johon syötetään videon minuuttimäärä. Ohjeistusta opiskelijoille uusista kentistä on lisätty syöttösivulle, ja tarvittaessa Theseus-asiantuntijat tekevät erillisen ohjeen.

3. Edellisen kokouksen muistio

Käytiin läpi edellisen kokouksen (22.5.2019) muistio. Todettiin, ettei muistio ole vielä wikissä. Jyrki Ilva lähettää kommenttinsa muistioon Minna Marjamaalle sähköpostilla.

Kiinnitettiin huomio seuraaviin muistion kohtiin.

4. Word-muotoisten tiedostojen lataamisen esto Theseukseen

- Word-tiedostojen määrä ei ole helposti selvitettävissä, joten päätettiin luopua selvityksen tekemisestä. Keskustelussa ilmeni, että jos Word-tiedostot poistetaan DSpace-alustan tiedostomuodoista, se vaikuttaa myös aiemmin talletettujen tiedostojen käsittelyyn, sillä selain ei osaa avata niitä.
- **Päätös:** Poistetaan Word-tuetuista tiedostomuodoista. Aiemmin tallennetut Word-dokumentit korjataan manuaalisesti sitä mukaa, kun niitä tulee ilmi. Samu kertoo Theseus-asiantuntijoille, kun muutos on tehty. Theseus-asiantuntijat ohjeistavat opiskelijoita.
- **Lisäys 8.10.19:** Samu on tutkinut asiaa kokouksen jälkeen. Word-tiedostoja on 33 kappaletta, ja Theseus-rukkaset ovat saaneet niistä listan. Jos tiedostot estetään, ne katoavat metatietoineen näkyvistä käyttäjiltä. Tiedostojen muuntaminen PDF-muotoon on yksi vaihtoehto.

5. Haka-federaation tietojen perusteella opiskelijan tallennuksen pakottaminen oman ammattikorkeakoulun opinnäytetyökokeelmaan

- Minna Marjamaa on kysynyt AAPA-verkostosta eduPersonAffiliation-attribuutin käyttöönotosta, mutta kaikki organisaatiot eivät ole vastanneet. Attribuuttia ei voi ottaa käyttöön, ellei se ole kaikilla käytössä.
- **Päätös:** Minna kysyy uudelleen verkostolaisilta.

7. Haaga-Helian KONTO

- Samu on maininnut REST-rajapinnan käyttömahdollisuudesta tässä yhteydessä Haaga-Heliale ja pyytänyt kuvausta siitä, miten järjestelmän olisi tarkoitus toimia Theseuksen kanssa yhteen, mutta sieltä ei ole sittemmin otettu yhteyttä.
- **Päätös:** Antti puhuu asiasta Eevan kanssa.

8. Saavutettavuusdirektiivi

- Theseuksen käyttöliittymän saavutettavuus on Kansalliskirjaston ja talletettujen aineistojen saavutettavuus organisaatioiden vastuulla. AMK-konsortion puolelta asia ei ole edennyt. Palvelulle tarvitaan saavutettavuusseloste, yksi seloste riittää. Ohjaus- ja valvontavastuu on Aluehallintovirasto AVIlla, joka on tehnyt työkalun saavutettavuusselosteen luomiseksi. Sisällöt vaikuttavat hankalimmalta osalta. Celia on tehnyt ohjeistusta tiedostojen saavutettavuudesta. Kansalliskirjaston asiakaswikiin on tulossa sivu, joka kokoaa Celian ohjeet yhteen.
- Samu kertoi kuulleen Turun yliopistolta, että Muuntajan[1] käyttö on ongelmallista. Muuntaja ei tuota saavutettavia PDF-tiedostoja kaikissa tapauksissa, koska joissain kohdin PDF/A-tiedostomuodon ja saavutettavuuden välillä on ristiriitoja.
- **Päätös:** Kansalliskirjasto edistää wikisivua, jolle lisätään myös AVIn työkalun linkki. Saavutettavuusseloste tehdään yhteistyössä.

10. Kansikuvan muodostumatta jääminen

- Tiedostoa vaihdettaessa kansikuva ei generoidu kuin manuaalisesti. Asiaa on selvitetty, ja käsin vaihtaminen on ainoa vaihtoehto. Tämä ei koske vain Theseusta.
- **Päätös:** Theseus-asiantuntijat ilmoittavat havaitsemistaan puuttuvista kansikuvista kansalliskirjastolaisille, jotka vaihtavat ne. Asiaan yritetään edelleen löytää parempi ratkaisu.

11. Justus-Theseus-integraation toteutus ja korjaukset

- Integraatio on avattu viime toukokuussa. Kansalliskirjastoon ei ole raportoitu ongelmista. CSC:llä yhteyshenkilö on Joonas Nikkanen. Antti kertoi ongelmista mm. menneisyydessä olevan embargoajan kohdalla. Ongelmat on kuitenkin korjattu, eikä uusia ole tullut.
- **Päätös:** Todettiin tilanne vakiintuneeksi.

12. OpenAire-kentät ja Core

- Theseuksen metadatauudistuksessa ei ole luotu OpenAire-metatietokenttiä. Theseus-asiantuntijat ovat olleet kiireisiä massaeditointien kanssa, joten asia ei ole edistynyt AMK-puolella. Uusin OpenAire-versio 4.0 ei ole käytettävissä nykyisessä DSpace-versiossa, sillä vasta tuleva DSpacen versio 7.0 mahdollistaa hierarkisen tietomallin. Coreen on mennyt Theseuksen opinnäytetöitä, vaikka niitä ei olisi sinne haluttu. Erityisen ongelmallista on Coren omalle palvelimelleen imuroimien pdf-tiedostojen julkinen näkyvyys. Jatkossa pitää selvittää, tarjoavatko Coren omat työkalut mahdollisuuksia rajata pdf-tiedostojen tallentumista. Suomessa Coren työkaluihin ei ole juurikaan tutustuttu, joten selvitystyö olisi tarpeen.
- **Päätös:** Asia jää tässä vaiheessa avoimeksi.

4. Theseuksen käyttökatkot syksyllä

Tänä syksynä tapahtuneet käyttökatkot ovat olleet erilaisia kuin aiemmin, ongelmia on ollut vain osalla käyttäjistä, minkä tähden niihin reagoitiin hieman hitaammin. Ongelmiin vaikutti klusterirakenne, jossa kaksi edustapalvelinta replikoi istuntoja keskenään. Jos kirjautuminen tapahtui ykköskoneella ja istunto siirtyi sen jälkeen kakkokoneelle, replikointi katkesi ip-pakettien salauksen toimintahäiriöön. Ongelman selvittämisenopeuteen vaikutti myös se, että järjestelmässä on nyt klusterirakenteesta ja Haka-kirjautumisesta johtuen useampi kohta, josta ongelmaa jouduttiin aluksi etsimään.

Ratkaisua etsittiin uudella konfiguroinnilla, mutta se ei toiminut, joten jouduttiin palaamaan edelliseen konfigurointiin. Uusi ratkaisu on löytynyt, klusteri siirretään suljettuun sisäverkkoon, mutta sen toteuttaminen vaatii aikaa ja onnistuu vasta ensi vuoden alussa tammi-helmikuussa, sillä työtä ei voi tehdä syksyn kiireisimpänä aikana. Sillä välin Theseus uudelleenkäynnistetään hallitusti kerran viikossa huoltoikkunassa. Todellinen katko käyttäjälle arvioidaan n. viiden minuutin mittaiseksi.

Päätös: Kerran viikossa tehdään hallittu käyttökatko Theseuksen uudelleenkäynnistämiseksi. Huoltoikkuna on maanantaisin klo 13-14, ja sen aikana tulee tyypillisesti yksi käyttökatko, jonka kesto on noin viisi minuuttia. Edustakoneelle laitetaan katkon ajaksi huoltokatkoilmoitus. Lisäksi huoltoikkunasta laitetaan varoitusteksti [submissions.theses.fi](https://theses.fi/submissions)-sivulle.

5. PAS-tilannekatsaus

Kesän aikana on kartoitettu mallia julkaisuarkistojen PAS-siirroille. Julkaisuarkistoaineistoa kohdellaan vapaakappaleina, mikä mahdollistaa aineistojen viemisen CSC:n PAS-järjestelmään Kansalliskirjaston vapaakappaleprosessin puitteissa. Kansalliskirjaston johto on pyytänyt ylimääräistä sisäistä selvitystä, joka valmistuu marraskuun alussa, ja käytännön toteutustyö voidaan aloittaa vasta selvityksen jälkeen.

Aikataulusta ei ole tarkkaa tietoa, se riippuu selvityksen tuloksesta. Toteutukseen osallistunee asiantuntijoita Kansalliskirjaston tutkimuskirjaston vapaakappaletoimistosta, jonka vastuulle PAS on organisaatiouudistuksen myötä siirtynyt, sekä Kirjastoverkkopalveluiden julkaisemisen palvelut -ryhmästä. Työnjako ei ole vielä selvillä.

Päätös: Seurataan tilannetta.

6. Sisältöjen nostot etusivulle

AMK:t haluaisivat itse voida vaikuttaa siihen, mitä AMK-kohtaisella etusivulla on näkyvissä, siten että sivulta pääsisi suoraan esim. tietyn oppiaineen opinnäytteisiin tai julkaisuihin. Samu esitteli Turun yliopiston väitöskirjakokoelmaa, jossa on linkit aineistoihin sarjoittain jaoteltuina. Tämän saa aikaan kirjoittamalla HTML-koodin kokoelman Description-kenttään, ja linkiksi voi määritellä minkä tahansa kenttähaun tuloksen.

Päätös: Samu, Antti, Tiina ja Minna M. aloittavat työn erillisellä verkkokokouksella ja jatkavat testaamalla omilla AMK:llaan. Minna M. tekee perjantaina Doodlen kokousajankohdan sopimiseksi.

7. Justuksen kehittäminen

Suurin osa ammattikorkeakouluista on ottanut Justuksen käyttöön, eikä integraatiossa ole ollut ongelmia. AMK-kentällä olisi hyvä olla yhteinen taho, jotta Justuksen kehittämiseen voitaisiin vaikuttaa. Kansalliskirjastoa kiinnostaa tietää, mikäli Justus aiheuttaa muutostarpeita Theseukseen. Todettiin, että metadata on keskeistä, ja uusia kenttiä on lisätty Theseukseen tarpeen mukaan.

Keskusteltiin myös yliopistojen käyttämästä Pure-järjestelmästä, jonka käyttöönotto on ollut esillä joissakin yliopiston yhteydessä toimivissa AMK:eissa. Jyrki huomautti, että Theseus-Pure-kytkennän toteuttaminen voi olla vaikeaa. AMK:t ovat suunnitelleet ensi vuodelle Justus-seminaaria CSC:n kanssa.

Päätös: Seurataan tilannetta.

8. Theseuksen ylläpito ja kehittäminen

Tarkasteltiin alustavaa toimintasuunnitelmaa vuodelle 2020. Kansalliskirjasto toivoi siihen lisäystä, että Kansalliskirjasto hallinnoi Theseuksen lisäksi useaa muutakin julkaisuarkistoa ja että niiden kehittäminen saattaa tuottaa parannuksia myös Theseukseen. Kehitteillä on mm. URN-pikarekisteröinti, joka voidaan ottaa käyttöön kaikissa arkistoissa.

Theseukseen Kansalliskirjastossa käytettäviä ensi vuoden resursseja on vaikea arvioida, sillä työtä tehdään osana laajempaa infrastruktuuria, ja on hankala sanoa, mikä osa työstä kohdistuu juuri Theseukseen. Theseus on kuitenkin iso asiakas, esimerkiksi klusteriratkaisu, massaeditointi, kokoelmarakenteen purku ja suuret aineistomäärät työllistivät merkittävästi helmi-kesäkuussa 2019.

Theseus-rukkasten resurssimitoitus ensi vuodelle on 3*0,2 henkilötyövuotta. Todettiin, että tänä keväänä allokoitu resurssivaraus on ylittynyt roimasti.

Samu kertoi tiedotusasiaina, että Kansalliskirjaston Vapaakappalepalveluissa on pilotoitu Wayback Machine -ohjelmaa, jolla saadaan aikaan snapshot verkkosivustosta. Tästä on ePookin kanssa suunnitteilla demo.

9. Kustannukset vuonna 2020

Vuosikokoukseen tarvitaan kustannusarvio. Suuria korotuspaineita ei ole. Tämän vuoden julkaisuarkistolaskutus on saatu valmiiksi elokuussa. Tritoniaan ja LUT-konserniin kuuluvilla AMK:eilla oli vanhentuneet laskutustiedot.

Päätös: Jyrki toimittaa kustannustiedot Hannalle 21.10. mennessä. Hanna pyytää Tritonia- ja LUT-ammattikorkeakouluja tarkistamaan laskutustietonsa.

10. Muita mahdollisia asioita

Jyrki kertoi DSpaceen version 7.0 viivästyneen. Tällä hetkellä tehdään arviota siitä, paljonko jäljellä oleva kehitystyö vie aikaa. Uudistustyö on suurelta osin tehty, mutta kyseessä on suuria ja vaikutuksiltaan laajoja muutoksia, ja työn viimeistely on kesken. On mahdollista, että betaversio julkistetaan vasta ensi vuonna, ja lopullinen tuotantoversio saatetaan julkistaa pian sen jälkeen.

Kansalliskirjasto on sitoutunut DSpaceen alustana eikä harkitse siirtymistä esimerkiksi Fedoraan. Kehitystyötä toki seurataan.

11. Seuraava laatupalaveri

Seuraava kokous on tiistaina 10.3.2020 klo 10-12.

12. Kokouksen päättäminen

Puheenjohtaja päätti kokouksen klo 14:50.

[1] <https://muuntaja.oulu.fi/>