



This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail.

Please cite the original version: Tolonen, T. ; Nyqvist, A. & Marjamaa, M. (2019) Iso laiva kääntyy hitaasti – missä mennään Theseus-uudistuksessa? Kreodi 2.

URL: <https://www.kreodi.fi/en/31/Artikkelit/555/Iso-laiva-k%C3%A4%C3%A4ntyy-hitaasti-%E2%80%93-miss%C3%A4-menn%C3%A4n-Theseus-uudistuksessa.htm>

Iso laiva kääntyy hitaasti – missä mennään Theseus-uudistuksessa?

Tiina Tolonen
Antti Nyqvist
Minna Marjamaa

Hartaasti odotettu ja suunniteltu Theseus-päivitys käynnistyi helmikuun lopussa. Kaksi viikkoa tehtiin pitkää päivää töitä, myös viikonloppuna. Katkon jälkeen töitä riitti rukkanpostiin tulleiden ongelmien ratkaisuisissa ja kysymyksiin vastailussa. Nyt iso osa työstä on takana. Mitä oikeastaan tapahtui ja mitä tästä kaikesta oli hyötyä?

Kuten edellisessä Kreodissa kerrottiin, luvassa oli Theseuksen metadatan päivitys, kokoelmarakenteen muuttaminen sekä syöttölomakkeiden uudistaminen. Kansallinen metadatasuositus julkaisuarkistojen tekstiaineistoille valmistui jo edellisvuonna, mutta koska päivitystä varten tarvittiin pitkähkö Theseus-katkos, päivitystyön aloittaminen siirtyi rauhalliseksi arvioituun alkuvuoden ajankohtaan. Ajankohdan valintaan vaikutti myös kohta avautuvan Justus-Theseus-integraation valmistuminen, sen toivottiin tapahtuvan myös samoihin aikoihin.

Syöttölomakkeen uudistus ja uusi kokoelmarakenne

Syöttölomakkeiden uudistamisen lisäksi tarkoitus oli ottaa käyttöön uusi lomaketekniikka, sillä vanha räätälöity syöttölomake oli teknisesti vanhentunut – vanhan lomakkeen muokkaaminen uuden metadataskeeman ja kokoelmarakenteen mukaiseksi olisi todennäköisesti ollut hyvin työläs prosessi. Samassa yhteydessä vanha shibboloimaton syöttölomake oli poistumassa käytöstä. Uusi syöttölomake on huomattavasti helpommin muokattavissa kuin vanha lomake ja mahdollinen Dspacea tapahtuva kehitystyö on joustavasti otettavissa siinä käyttöön.

Kokoelmarakenteen muutoksen lähtökohtana oli se, että Justus-integraation vuoksi julkaisukokoelmien määrä per ammattikorkeakoulu tuli supistaa yhteen. Samassa yhteydessä päätettiin tehdä sama myös opinnäytekkokoelmien kanssa ja luopua jatkuvasti paisuvasta kokoelmavalikoimasta. Metadatan osalta tärkeintä oli varmistaa se, että Theseus noudattaa kansallista metadatasuositusta. Sitä tarvitaan nykyisiä ja tulevia toiminnallisuuksia varten.



Kolmesta palvelimesta kuuteen

Myös Theseuksen teknisellä puolella tapahtui uudistuksia. Aikaisemmin Theseus-palvelinkokonaisuus koostui kolmesta palvelimesta eli Theseus.fi sivua pyörittävästä palvelimesta, opinnäytetöiden syöttöön tarkoitusta palvelimesta, jossa oli käytössä Kansalliskirjastolla alusta asti tehty sovellus syöttämistä varten sekä tallennuksen aloitusta ja ohjeistusta varten tarkoitettua palvelimesta.

Nyt Theseus koostuu kuudesta palvelimesta, ainoastaan syöttöön tarkoitettu palvelin on pysynyt ennallaan. Balancer-edustapalvelin vastaa kaikkiin Theseus.fi-sivustolle tuleviin pyyntöihin, Dspacen tietokanta on siirretty omalle palvelimelleen ja SOLR-indeksointia varten on myös oma palvelimensa. Käyttökuorma on siis hajautettu sekä applikaation, tietokannan, hakuindeksin, että tiedostojen jakelun osalta eri palvelimille. Vaikutukset ovat havaittavissa mm. siten, että Theseuksen vasteajat ovat nyt huomattavasti nopeampia kuin aikaisemmin.

Uudistusten takana on Theseuksen suuri käyttäjämäärä ja Dspacen oman syöttöjärjestelmän käyttöön ottaminen, joka lisää käyttökuormaa edelleen. Syöttölomakkeen on valmistuvien opiskelijoiden kannalta oltava toiminnassa, vaikka huoltoja tehtäisiin. Kaksi applikaatiopalvelinta mahdollistaa tämän, eli toinen voi olla huollossa ja toinen palvelee edelleen syöttäjiä / tiedonhakijoita.

Valmistelutyö ennen katkoa

Valmistelutyö käynnistyi syksyllä 2018 metadatasuunnitelman läpikäymisellä. Alustava suunnitelma oli jo olemassa, mutta sitä oli tarvetta tarkistaa erityisesti eri amkeista tulleiden kenttäehdotusten ja -toivomusten pohjalta. Tässä yhteydessä käytiin läpi myös Theseuksen tämänhetkinen metadata, muuttuneet kentät ja niiden arvot. Näiden läpikäyminen on osaltaan jatkunut aivan näihin hetkiin saakka, kun kokoelmamuutoksen vaatimia massaeditointeja on tehty. Vaikka Theseus onkin ollut aina metadataaltaan hyvin yhtenäinen, se ei poista sitä mahdollisuutta että etteikö amkeissa on käytetty samaa kenttää eri tarkoitukseen tai tehty kokoelmien sisäisiä mappauksia niin, että ne pistivät massaeditoinnin umpisolmuun.

Kirjautumisen ongelmat katkon jälkeen

Vaikka ensi alkuun vaikutti siltä, että kahden viikon katkos olisi riittävän pitkä, se osoittautui aivan liian lyhyeksi monestakin eri syystä. Theseus avattiin uudelleen 13.3., jolloin heti ilmeni ongelmia Shibbolethin uudelleenohjauksen kanssa. Syöttösivun linkkejä ei saatu toimimaan oikein siten, että kirjautumisruudun jälkeen tallentaja olisi siirtynyt suoraan tallennussivulle. Ongelman vuoksi ruotsin- ja englanninkieliset sivustot olivat poissa käytöstä pidempään, kun Kansalliskirjastossa ratkottiin asiaa ensin yhdellä kielellä ratkaisun löytämisen nopeuttamiseksi sekä ylimääräisen sekaannuksen välttämiseksi. Tästä seurasi myös ajoittaisia virheilmoituksia sekä syötössä että tallennusten editoinnissa.

Katkon aikana ei kaikkia massaeditointeja saatu tehtyä valmiiksi ja niitä jatketaan edelleen. Massaeditoinnin tarve osoittautui arvioitua suuremmaksi, koska uuden metadataskeeman myötä tuli mahdolliseksi nostaa tallennettuja tietoja eri tavalla esille etusivun selausfasetteihin. Tällainen on esimerkiksi opinnäytetyön taso, jonka avulla pystyy nyt selaamaan vaikkapa kaikkia YAMK-opinnäytetöitä yhtä aikaa tai ammattikorkeakouluittain.

Yllätyksiä

Vaikka Theseus-laatupalaverissa yhdessä Kansalliskirjaston kanssa oli käyty läpi Theseukseen päivityksen yhteydessä tulevia muutoksia sekä positiiviselta että negatiiviselta kannalta, pääsivät muutamat asiat yllättämään molemmat tahot. Tällainen oli esimerkiksi opiskelijalle heti tallentamisen jälkeen aiemmin lähtenyt vahvistusviesti, joka nyt lähtikin vasta kun työ oli käsitelty kirjastossa. Samaten viestistä puuttunut URN-tunnus herätti kysymyksiä. Väärään kokoelmaan tallennettua työtä ei pysty enää siirtämään kokoelmasta toiseen, kuten aiemmin. Tämäkin tuli tietoomme vasta siinä vaiheessa, kun ensimmäinen virheellinen tallennus tehtiin.

Onneksi mitään katastrofaalista ei kuitenkaan tapahtunut, vaikka ensimmäisinä päivinä katkon päättymisen jälkeen Theseus-rukkasten sähköpostiin vyörynyt palautteen määrä olikin huimaava ja aiheutti lievää paniikin tunnetta. Kaikki palaute otetaan vakavasti ja selvitetään onko asialle mahdollista tehdä jotakin ja missä aikataulussa. Edelleen toivomme kärsivällisyyttä, vaikka nyt saattaakin näyttää, että uudistus oli askel huonompaan suuntaan. Sen hienous ja vaikuttavuus Theseuksen käyttöön konkretisoituu vasta siinä vaiheessa, kun massaeditoinnit on saatu loppuun. Kun kaikki oleellinen tieto on löydettävissä metadatatista, paranee sen uudelleenkäyttö ja eri järjestelmät pystyvät hyödyntämään Theseuksen dataa tehokkaammin.

Vastauksia kysymyksiin

Theseuksen SimpleStats-tilastot ovat katkon jälkeen olleet poissa käytöstä. Tilastoja kertyy normaalisti, mutta raportointiohjelma vaatii klusterointiympäristön vuoksi uutta koodia ja sen teko on kesken. Tilastoihin liittyvä työ on vielä kesken, sillä samalla on muutettu takautuvien tietojen esitystapaa Simplestatsin omassa käyttöliittymässä uuden kokoelmarakenteen mukaiseksi, mikä on ollut hieman isompi projekti. Uusimmat käyttö tiedot eivät tämän takia vielä tule näkyviin item-näytössä, vaan siellä näkyvät latausluvut perustuvat toistaiseksi takautuviin tietoihin.

Opiskelijalle lähetettävä automaattinen vastausviesti on työn alla, samoin mahdollisuus saada siihen tallennetun työn URN.

Theseuksesta on edelleen mahdollista saada RSS-feediä. RSS- ja ATOM-syötteitä voi tehdä hakuihin perustuen OpenSearchin avulla, Kansalliskirjaston sivuilla kerrotaan lisää asiasta. <https://www.kiwi.fi/pages/viewpage.action?pageId=45782169>

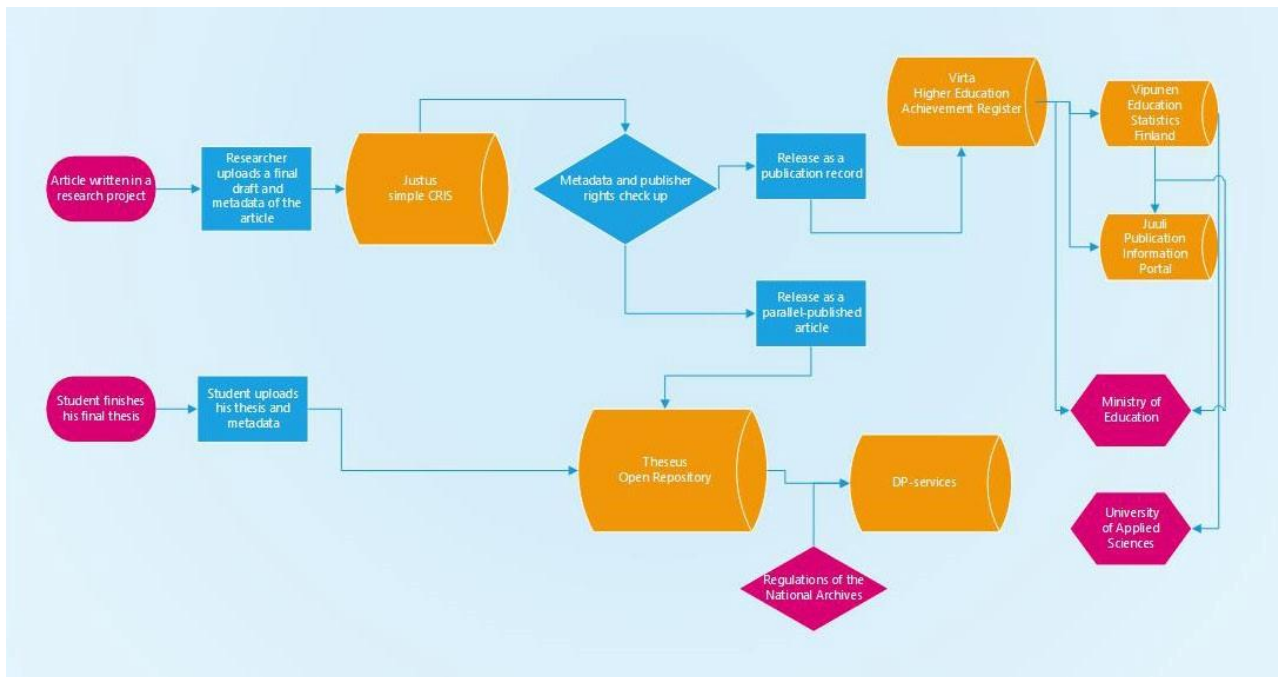
Jos opiskelija tallentaa työn muussa kuin PDF-muodossa, kansikuvaa ei silloin generoidu. Tässä tapauksessa henkilökunnan kannattaa muuttaa työ PDF-muotoon ja valmistaa kannesta jpg-muotoinen kansilehti ja tallentaa se Theseukseen nippuun Esikatselukuvat.

Viittausboxi julkaisuille on otettu uudestaan käyttöön. Formaatti viittausboxille tulee kansallisen työryhmän luomasta formaatista. Jos boxin tiedoissa on ongelmia, se johtuu metadatatassa olleista ongelmista eli korjaamalla metadattaa saa viittausboxia korjattua.

Theseukseen mahdollisesti tarvittavia saavutettavuusdirektiivin mukaisia muutoksia selvitetään tänä vuonna niin ammattikorkeakoulujen käytänteinä kuin teknisen toteutuksen vaatimuksina. Ennen 23.9.2018 julkaistujen verkkosivustojen pitää olla saavutettavuusvaatimusten mukaisia 23.9.2020.

Kehitystyö jatkuu

Justus–Theseus-integraatio on testauksessa ja valmistunee parin viikon kuluessa. Työtä on viivästyttänyt ammattikorkeakoulujen 28.3. toivotat parannukset ennen palvelun avaamista.



Kaaviokuva Theseuksen pääprosesseista: julkaisun ja opinnäytetyön tallennuksesta ja niiden kytkeytymisestä Virtaan ja PASiin.

Suuri työ on vielä koulutus/tutkinto-ohjelmiin perustuneen kokoelmarakenteen muuttamisessa vanhojen tallennusten osalta sellaiseksi, että se vastaa uuden syöttölomakkeen koulutusala-vetoalikkaa. Tarkoituksena on mahdollistaa opinnäytetöiden selaaminen kattavasti ilman, että tietää miten kyseinen koulutus/tutkinto-ohjelma on nimetty kussakin ammattikorkeakoulussa. Toistaiseksi tutkinto-ohjelmien massaeditointeja on tehty vasta viiden ammattikorkeakoulun kokoelmiin.

Theseuksen ammattikorkeakoulujen omia etusivuja on myös tarkoitus muokata ammattikorkeakoulun toiveita vastaamaan niin, että amkit saavat esille toivomiaan aineistoja. Tämä työ aloitetaan vasta tutkinto-ohjelmien massaeditointejen jälkeen.

PAS-käytänteiden selvitys jatkuu vielä tänä keväänä, samoin streamauksen käyttöönotto. Tiedotamme näistä, kun työ etenee.

AMKIT-konsortion kirjastonjohtajien kokouksen päätöksellä Theseus-asiantuntijaksi on valittu kolmas henkilö, Haaga-Helian Antti Nyqvist. Toivotamme Antin tervetulleeksi remmiin!

Tietoa Theseus-asiantuntijoista:

Tiina Tolonen, palvelupäällikkö Oulun ammattikorkeakoulu

Antti Nyqvist, tietopalvelusihteeri Haaga-Helia

Minna Marjamaa, informaatikko Laurea