

E-thesis-palvelun toimintakertomus 2004-2005

E-thesiksen toimintaan on ollut vuosien 2004-2005 aikana suunnitteilla merkittäviä teknisiä uudistuksia, jotka alkavat näkyä julkaisijoille ja palvelun käyttäjille vuoden 2006 alussa. Uudistusten suunnittelu on ollut läheisesti kytköksissä alkuvuodesta 2005 käynnistettyyn väitöstietohankkeeseen, jonka myötä Helsingin yliopiston väitöstietojen keruu ja väitöstietojen julkaiseminen keskitetään E-thesiksen yhteyteen.

1. Toiminnan resurssit

E-thesis-palvelulla oli helmikuusta 2004 lähtien käytössään 1,4 htv:tä (1,0 x atk-suunnittelija ja 0,4 x suunnittelija/sovellussuunnittelija). Väitöstietohankkeen tarpeisiin saatiin lisäksi yliopistolta 0,6 htv:tä vuosiksi 2005 ja 2006, mikä mahdollisti uuden työntekijän palkkaamisen toukokuussa 2005. E-thesiksen kokonaisresurssit olivat siis yhteensä 2,0 henkilötyövuotta: näistä 1,6 htv:tä saatiin yliopistolta ja 0,4 htv:tä HYK:in budjetista.

Käytännössä E-thesiksen toiminnasta ovat vastanneet sovellussuunnittelija Jyrki Ilva, atk-suunnittelija Jani Saijos (helmikuusta 2004 lähtien) ja atk-suunnittelija Jussi Piipponen (toukokuusta 2005 lähtien). Syksyllä 2005 työnjako oli sellainen, että Ilva vastasi pääosin palvelun hallinnosta, tilastoinnista ja yhteyksistä sidosryhmiin, Saijoksen vastuualueena oli tekninen kehittämistyö ja Piipponen puolestaan huolehti julkaisutoimintaan liittyvistä käytännön töistä ja asiakaspalvelusta.

2. Julkaisutoiminta

E-thesis on jatkanut Helsingin yliopiston opinnäytteiden ja sarjajulkaisujen julkaisemista aiemmin luodun toimintamallin mukaisesti. Väitöskirjat ovat säilyttäneet asemansa tärkeimpänä julkaistavana materiaalina, vaikka myös muiden aineistojen määrä on kasvanut jonkin verran. Yliopiston kaikista väitöskirjoista E-thesiksen kautta ilmestyi noin kolme neljäsosaa, muissa opinnäytteissä elektronisten julkaisujen osuus oli edelleen huomattavasti pienempi.

Vuonna 2004 E-thesiksessä ilmestyi yhteensä 471 julkaisua, joka oli siihenastinen ennätys. Näistä 322 oli väitöskirjoja (v. 2003: 237), lisensointitoita julkaistiin kolme (v. 2003: 3) ja graduja 72 (v. 2003: 35). Sarjajulkaisuja ilmestyi 74 (v. 2003: 37). Vuoden 2005 luvut eivät ole vielä valmiit, mutta vuoden kolmen ensimmäisen neljänneksen perusteella näyttää siltä, että

uusien julkaisujen määrä tulee olemaan edellisvuoden tasolla. Kaikkiaan E-thesiksessä on vuoden 2005 loppuun mennessä ilmestynyt jo yli 2000 julkaisua.

Yliopistopainon markkinaosuus väitöskirjojen tiedostojen toimittajana on pysynyt suhteellisen vakaana, noin 40-50% välillä. Muiden julkaisujen materiaali tuli pääosin suoraan tekijöiltä itseltään. Lisäksi Helsingin yliopiston kirjasto ja Valtion teknillinen tutkimuslaitos sopivat joulukuussa 2004 siitä, että E-thesis saa julkaistavakseen myös VTT:n julkaisusarjassa ilmestyvät Helsingin yliopiston väitöskirjat, jotka olivat aiemmin sopimusteknisistä syistä johtuen ilmestyneet vain VTT:n omilla verkkosivuilla. Näiden väitöskirjojen osalta VTT huolehtii julkaisuluvan hankkimisesta ja tiedostojen toimittamisesta E-thesikseen.

3. Julkaisujen käyttö

E-thesiksen julkaisujen yhteenlaskettu käyttö on kasvanut sitä mukaa kuin julkaisujen määrä on lisääntynyt. Koko E-thesis-palvelimen tasolla kehitys näkyy siten, että palvelimelta ladatun datan yhteismäärä on noin kaksinkertaistunut syksyn 2003 ja syksyn 2005 välillä, samalla kun julkaisujen yhteismäärä on kasvanut vähän yli tuhannesta noin kahteen tuhanteen. E-thesiksen julkaisujen yhteenlasketuista latauskerroista on kattavia lukuja vasta vuodelta 2004, jolloin latauskertojen kokonaismäärä ylsi hieman yli puoleen miljoonaan, vaikka luvuista vähennetään hakurobottien aiheuttama käyttö.

Yksittäisten julkaisujen kohdalla latauskertojen määrä näyttää sen sijaan aiempiin vuosiin verrattuna pysyvän lähes vakiona. Enimmillään julkaisut yltyvät kuukausittain muutamaan sataan latauskertaan, keskiarvo taas on noin 30 latauskertaa kuukaudessa. Kaikkien aikojen ennätys koettiin syyskuussa 2004, kun Taina Myöhäsen folkloristiikan pro gradu –työ "Nimettömät näkyvillä" keräsi kuukauden aikana peräti 4315 kävijää. Kunkin kuukauden käytetyimmät julkaisut jakautuvat selkeästi kahteen eri ryhmään: toisaalta suosituimpien joukkoon yltää tietty joukko kestosuosikkeja, joita ladataan tasaisesti kuukaudesta toiseen, toisaalta suosituimpien joukkoon yltää myös säännöllisesti joitakin uusia julkaisuja, jotka ovat esillä E-thesiksen omilla uutuussivuilla tai laajemminkin mediassa.

Tilastojen mukaan vain noin 9% julkaisujen käytöstä tulee Helsingin yliopiston omista verkko-osoitteista, eli ne tavoittavat runsaasti myös yliopiston ulkopuolisia lukijoita. Kotimaisten käyttäjien osuus on verkko-osoitteista päätellen hieman yli puolet. E-thesiksen käyttötilastojen mukaan suurin osa käyttäjistä löytää julkaisut joko Google-hakukoneen tai E-thesiksen omien verkkosivujen kautta. Uutuusjulkaisujen aktiiviseen markkinointiin E-thesis-palvelulla ei ole itsellään ollut resursseja muuten kuin joissakin poikkeustapauksissa.

4. Kehittämishankkeet

4.1. E-thesiksen siirtäminen Doriaan

E-thesiksen teknisen toimintaympäristön kehittämistä ryhdyttiin suunnittelemaan jo syksyllä 2003. E-thesiksellä ei ole tähän asti ole ollut käytössään omaa tietokantaratkaisua, vaan sen julkaisut ja niiden metatiedot on sijoitettu tavanomaiselle verkkopalvelimelle. HYK:in tuolloin hankintavaiheessa ollut digitaalisten aineistojen hallintajärjestelmä näytti tarjoavan myös E-thesiksen kannalta toimivan ratkaisun.

Tarjouskilpailun myötä valitun ENCompass-ohjelmiston käyttöönotto viivästyi useista eri syistä johtuen niin, että se saatiin testikäyttöön vasta alkusyksystä 2004. Ensimmäinen tämän Doriaksi nimetyn järjestelmän avulla tuotettu kokoelma julkistettiin huhtikuussa 2005, ja myös E-thesiksen oma Doria-kokoelma on ollut alustavassa testikäytössä kevästä 2005 lähtien. E-thesiksen väitöstietohankkeen myötä laajentuneiden tehtävien vuoksi uuden Doria-käyttöliittymän lopullista käyttöönottoa on kuitenkin päätetty lykätä siihen asti, kunnes uusi metatietojen keruuseen käytettävä järjestelmä on kokonaisuudessaan valmis.

Uusien julkaisujen lisäksi Doriaan täytyy siirtää myös E-thesiksen tähänastisten noin 2000 julkaisun metatiedot, joita ei ole aiemmin tallennettu tietokantaan. Tätä tarkoitusta varten keväällä 2005 aloitettiin erityisen metatietojen keruuseen soveltuvan parseriohjelman suunnittelu. Ohjelman avulla vanhojen julkaisujen metatiedot voidaan kerätä suurelta osin automaattisesti E-thesiksen www-sivuilta ja muuntaa Dorian kannalta soveltuvaan XML-muotoon.

4.2. Väitöstietohanke

Helsingin yliopiston kirjastotoimikunta asetti keväällä 2004 työryhmän "valmistelemaan väitöskirjoja koskevan tiedotuksen kehittämistä ja erityisesti tiivistelmien hyödyntämistä". Työryhmä päätyi esittämään, että yliopiston väitöstietojen keruu ja tiivistelmien julkaiseminen keskitettäisiin tulevaisuudessa E-thesis-palvelun yhteyteen. Väitöstietotyöryhmän suosituksen seurauksena Helsingin yliopiston kirjasto sai rahoitusta vuosille 2005-2006 väitöstietojen keruuseen ja tallentamiseen soveltuvan järjestelmän rakentamiseen ja käyttöönottoon. Käytännössä hankkeen vaatimasta teknisestä kehitystyöstä on vastannut pääosin atk-suunnittelija Jani Saijos. Yhteistyöllä tietokantapalveluiden muiden suunnittelijoiden kanssa on tosin myös ollut suuri merkitys hankkeen etenemiselle.

Väitöstietohankkeen ensimmäinen tavoite oli väitöstietojen ja tiivistelmien tallennukseen sopivan metadataformaatin luominen. Tietojen vaihdettavuuden takaamiseksi Dublin Core – pohjainen formaatti pyrittiin laatimaan sellaiseksi, että sitä voitaisiin soveltaa myös kansallisella tasolla, sekä Doriaa hyödyntävissä oppilaitoksissa että muissakin teknisissä järjestelmissä. Vaikka eri yliopistojen intressien yhteensovittaminen osoittautui ennakoituakin mutkikkaammaksi, formaatista saatiin aikaan esittelykelpoinen versio toukokuuhun 2005 mennessä.¹ Helsingin yliopiston lisäksi se on jo kesän kuluessa otettu käyttöön muutamassa muussakin korkeakoulussa.

Hankkeen toinen keskeinen päämäärä on väitöstietojen keruuseen ja käsittelyyn soveltuvan verkkolomakkeen luominen. Suunnittelun pohjana käytettiin E-thesiksen aiempaa, verkkojulkaisujen nimiösivun luomiseen käytettyä lomaketta, joka kuitenkin koodattiin alusta lähtien uudelleen. Tarkoitus on, että lomakkeen avulla kootuista tiedoista tulostetaan automaattisesti erilaisia tiedostoja eri tarkoituksia varten. Yksi tiedostoista on valmiiksi opinnäytteiden metadataformaatin mukaisessa XML-muodossa, jolloin se voidaan tietojen tarkistamisen jälkeen siirtää Doriaan. Lisäksi järjestelmä tuottaa automaattisesti yliopiston viestintäosaston tarvitsemat tiedot.²

Väitöstietohanke viivästy j jonkin verran alkuperäisissä suunnitelmissa kaavaillusta aikataulusta. Hanke-esityksessä oletettiin näet, että väitöstietolomake olisi ollut testattavana jo toukokuussa 2005 ja käytössä syyslukukauden alussa. Todellisuudessa testivaiheeseen asti päästiin vasta elo-syyskuussa 2005 ja käyttöönotto näyttäisi ajoittuvan vuodenvaihteeseen 2005/2006. Viivästyminen johtui suurelta osin siitä, että hanketta jouduttiin toteuttamaan sitä varten haettua 1,0 htv:tä pienemmillä henkilöresursseilla, mikä käytännössä tarkoitti sitä, että tarvittavan lisähenkilön palkkaaminen viivästy keväeseen alun perin suunnitellun alkuvuoden sijaan. Lisäksi lomakkeen kannalta oleellisen opinnäytteiden metadataformaatin suunnittelu kesti odotettua pidempään, osin muilta yliopistoilta saadusta palautteesta johtuen.

¹ Metadataformaatin tuorein versio on saatavilla E-thesiksen palvelimella osoitteessa <http://ethesis.helsinki.fi/metadata/>

² Lomakkeen testiversio on kokeiltavissa osoitteessa <http://ethesis.helsinki.fi/forms/diss.html>