

Open Repositories 2008

Southampton 1.-4.4.2008

Pekka Olsbo

Pertti Pitkänen

Matkasimme huhtikuun alkajaisiksi Southamptoniin, Etelä-Englantiin osallistumaan [Open Repositories 2008](#)-konferenssiin, maailman suurimpaan avointen julkaisuarkistojen kehittäjien, tutkijoiden ja visionäärien tapaamiseen. Päivitimme ajan tasalle tietojamme avointen julkaisuarkistojen kansainvälisestä kehityksestä ja tulevaisuuden suuntauksista sekä esittelimme tiistai-illan posterisessiossa Julkaisukanavahankkeessa tuotettua ratkaisua sopimustenhallinnan liittämistä julkaisuarkistoihin.

Posteresityksemme lisäksi päivillä oli paljon muutakin mielenkiintoista seurattavaa. Tiistai ja keskiviikko oli omistettu yleisille sähköisten julkaisuarkistojen kehittämiseen liittyville esityksille. Torstain ja perjantain esitykset oli jaettu julkaisuarkistosovelluksen mukaan erillisiksi tilaisuuksiksi [Dspacen](#), [Fedoran](#) ja [Eprinsin](#) käyttäjille. Erityisellä kiinnostuksella osallistuimme DSpace-ohjelmiston kehittäjien ja käyttäjien kokoukseen ([DSpace User Group Meeting](#)), jossa esiteltiin ohjelmiston uusin vasta julkaistu versio. Yliopistossamme käytössä oleva [JYX](#)-arkisto perustuu esitellyn sovelluksen edelliseen julkaisuversioon.

Konferenssin ohjelma löytyy konferenssin nettisivuilta. Lisäksi esitysmateriaaleille on perustettu oma Eprints-pohjainen [julkaisuarkisto](#), josta kaikki esitykset ja niiden materiaalit löytyvät. Lisäksi konferenssista on tuotettu esittelyvideo, joka löytyy Zepler.tv:n [sivuilta](#).

Tiistai & keskiviikko

Yhteisöllisyys – sitouttaminen – automatisointi

Konferenssin alkuosan teemoista vahvimmin nousivat esille arkistojen nykykehitykseen ja tulevaisuuteen liittyvät sosiaalisen verkon termit yhteisöllisyys (social networking), web2.0, annotointi ja yhteentoimivuus. Kuten moderneissa verkkopalveluissa yleensäkin, myös julkaisuarkistoissa käyttäjän rooli kasvaa tiedon hakijasta kohti tiedon kommentoijaa, käsittelijää ja tuottajaa. Monissa esityksissä käytettiin vertailukohtina suosituimpia uuden verkon palveluja kuten [Flickr](#), [iGoogle](#), [Picasa](#), Blogit ja Wikit.

Julkaisujen tallentamisen ja jakelun lisäksi myös tutkimustiedon säilöminen julkaisuarkistoihin tai tietoarkistoihin (Data Repositories) nähtiin avointen arkistojen käytön sovellukseksi. Aiheeseen pureutui mm. Peter Murray-Rust keynote-puheenvuorossaan [Repositories for Scientific Data](#).

Erilaisten järjestelmien yhteentoimivuus ja rajapinnat tulevat tärkeään asemaan, kun tiedon automaattinen siirtäminen palvelusta toiseen ja sen koostaminen käyttäjän tarvitsemiksi kokonaisuuksiksi on pystyttävä hoitamaan automaattisesti. Järjestelmien välistä yhteistoimintaa käsiteltiin tiistaina kolme esitystä käsittävässä osiossa nimeltä [Interoperability](#).

Julkaisuarkistojen kehityksen kannalta nähtiin olennaiseksi käyttäjän sitouttaminen palvelun käyttöön. Käyttäjällä tarkoitetaan arkiston näkökulmasta niin tiedon julkaisijaa kuin sen hakijaa. Julkaisijan osalta sitouttaminen voidaan toteuttaa integroimalla julkaisuutoiminnallisuus tiiviisti julkaisun tuottamisvaiheeseen ja työkaluihin, joilla julkaisua tuotetaan. Hakijalle on tärkeää tarjota varsinaisen tiedon lisäksi myös mahdollisuuksia tiedon kommentointiin, metatietojen rikastamiseen ja mahdollisesti jopa sen muokkaamiseen. Julkaisuarkistojen interaktiivista käyttöä ja hakijan kasvavaa roolia käsiteltiin tiistain kolmessa esityksessä, joiden ryhmä oli ytimekkäästi nimetty [Web 2.0](#). Samasta aiheesta jatkoi myös seuraava kolmen esityksen sessio [Social Networking](#).

Keskiviikkona päivän avasi lähinnä brittien kansallisia kehitysnäkymiä kartoittava kolmen esityksen sarja [National Perspectives](#). Koska Briteissä kehitetty Eprints-arkistosovellus on käytännössä saarivaltion akatemisen verkkojulkaisemisen de facto -julkaisualusta, esitykset käsitelivät sen ympärille rakennettuja sovellutuksia.

Seuraavaksi mielenkiintomme kohdistui arkistoihin liittyviä tekijänoikeuskysymyksiä käsittelevään esityssarjaan [Legal](#). Kuten jo alustuspuheenvuorossa mainittiin, esitykset jäivät varsin pinnallisiksi ongelmien esittelyksi ratkaisujen jäädessä lähes tyystin puuttumaan. Lisäksi esitykset keskittyivät oikeuksien määrittämisen rajoittavaan luonteeseen, eikä niissä huomioitu kaikkea sitä oikeuksien määrittämisen laajaa kirjoa, jota esimerkiksi hyvällä sopimustenhallintajärjestelmällä voitaisi kattaa. Näkökulma kohdistettiin julkaisun tekijään, eikä sopimustenhallinnan merkityksestä toiselle sopivalle osapuolelle, julkaisevalle yliopistolle, ollut juurikaan mainintaa. Session viimeinen, sinänsä varsin löyhästi edellisiin liittynyt esitys julkaisujen versionhallinnasta vei problematisoinnin huippuunsa.

Päivän viimeinen sessio [Usage](#) käsitteli arkistojen käytön analysointia ja käytöstä saatavan tiedon hyödyntämistä arkistojen kehittämisessä. Arkistojen käyttäjien käyttäytymisen tunteminen on avainasemassa kehitettäessä sovelluksia vastaamaan yhä paremmin käyttäjän tarpeita. Kun tiedetään, mitä käyttäjä hakee, millaisesta sisällöstä hän on kiinnostunut ja millaisia ”tämän julkaisun hakija kiinnostui myös tästä toisesta julkaisusta”-tyyppisiä suhteiden verkostoja sisältöjen välille voidaan rakentaa, voidaan paremmin ennakoida käyttäjän tiedontarvetta ja tarjota sopivaa sisältöä. Työkaluja arkistojen käytön analysointiin esiteltiin seuraavan päivän Repository Analytics –osoissa.

Torstai & perjantai

[Dspace user group meeting](#)

Seminaarin kaksi viimeistä päivää käytettiin eri julkaisuarkistosovellusten (Dspace, Fedora, Eprints) esittelyyn ja käyttäjäryhmien kokoontumiseen. Osallistuimme luonnollisesti pääosin Dspacea koskeviin tilaisuuksiin, joten raportin loppuosa käsittelee pääosin Dspace-kehitykseen liittyviä uutisia.

[Dspace 1.5 launch & Dspace future](#)

Torstai-aamuna Dspace foundationin johtaja Michelle Kimpton kertoi uusimmat järjestön tuoreimmat kuulumiset. Suurin uutinen oli tietenkin Dspace-ohjelmiston seuraavan version 1.5 julkaisu. Versiota luonnehditaan ohjelmiston historian tärkeimmäksi päivitykseksi. Myös yhteisössä on tapahtunut viimeisen puolen vuoden aikana. Lokakuun Rooman kokouksen aikaan järjestön organisoituminen oli vasta alkuvaiheessa eikä sillä ollut muuta henkilökuntaa kuin johtaja. Silloin avoimina olleet paikat on nyt täytetty. Yhteisöön on palkattu mm. teknologiajohtaja, jonka tehtävänä on koordinoita Dspace-ohjelmiston kehityssuunnitelmaa (roadmap for Dspace 2.0)

Case studies and best practises

Kahvitaun jälkeen kuulumme kolmen Dspacea käyttävän organisaation kokemuksia ohjelmiston käyttöön otosta ja räätälöinnistä.

eJamaica.org-portaali oli todella vaikuttava kokemus. Jamaikalla on rakennettu Dspacen pohjalle kansallinen julkaisufoorumi, joka pyrkii niin vapaaseen julkaisemiseen kuin mahdollista. Mielenkiintoisinta hankkeessa oli näkökulma Dspacen kokoelmarakenteeseen. Perinteisessä ajatusmallissa kokoelmat hahmotetaan ja suunnitellaan materiaalin ominaisuuksien mukaan. eJamaica-projektissa kokoelmat hahmotettiin käyttäjien kautta. Jokaiselle kirjautuneelle käyttäjälle luotiin automaattisesti oma kokoelma, jonka materiaaleja hän voi hallita täysin vapaasti. Käyttäjä voi mm. julkaista ja poistaa aineistoa, perustaa uusia alakokoelmia ja valtuuttaa muita käyttäjiä materiaalien käyttöön. eJamaica.org:n tavoitteena on kerätä kansalaisilta Jamaikan kulttuuriperintöön liittyvää informaatiota ja dokumentteja. Tällainen kansallinen palvelu ja kansalaisten aktiivinen osallistuminen kulttuuriperinnön säilyttämiseen lienee ainutlaatuista maailmassa.

Pandektis on Kreikkalainen hanke, joka keskittyi eri lähteissä olevien aineistojen siirtämiseen yhteen Dspace-arkistoon, materiaalikohtaisesti räätälöityihin näkymiin ja materiaalin linkittämiseen. Esityksessä kävi ilmi mitä kaikkia ongelmia puutteellisesti ja epäyhtenäisesti kuvailun aineiston siirtämisessä tulee vastaan.

Repository Analytics (Eprints session)

Itäpäivällä oli vuorossa julkaisuarkistojen käytön analysointiin kehitettyjen työkalujen esittely Sessiossa esiteltiin seuraavat tilastointityökalut ja –palvelut:

- IRStats, [AWStats](#): Logitietojen analysointiin kehitettyjä työkaluja. Näistä edellinen toimii vain Eprintsin kanssa.
- [Google Analytics](#): Googlen www-pohjainen palvelu, jonne arkiston voi rekisteröidä tilastotietojen keräämistä varten

Esittelyä vaille jäi Portugalilainen Minhon yliopiston kehittämä [työkalu](#), joka on kehitetty erityisesti DSpacen käytön analysointiin. Tosin sovellus toimii tällä hetkellä vain vanhan jsp-käyttöliittymän kanssa, joten Manakinia käyttävät arkistot eivät pysty hyödyntämään sen toimintoja ainakaan kovin helposti.

Dspace 1.5 training

Käyttäjryhmäkokouksen mielenkiintoisin anti sijoittui torstain viimeiseen osioon, jossa Dspacen kehittäjät havainnollistivat käytännön esimerkein uuden version

ominaisuuksia. Tilaisuus päättyi yli puoli tuntia kestäneeseen kysymystulvaan. Kysymysten määrä ei suinkaan kuvastanut uuden version julkaisemisen epäonnistumista vaan käyttäjäyhteisön valtavaa kiinnostusta uusiin ominaisuuksiin. Seuraavassa muutamia meidänkin kannaltamme mielenkiintoisia uusia toiminnallisuuksia.

Ohjelmiston lähdekoodin rakenne on laitettu aivan uuteen uskoon. Näkyvin esimerkki tästä on ohjelmiston asennustyökalun vaihtuminen Apache [Ant](#):sta [Maven](#):iin. Mavenin käyttö mahdollistaa lähdekoodin entistä paremman modularisoinnin. Jatkossa DSpace-ohjelmiston ytimen koodin määrä tulee olemaan entistä pienempi ja ohjelmisto kootaan liittämällä ytimeen haluttuja toiminnallisuuksia kuten webikäyttöliittymä ja muita rajapintoja. Julkaisuarkistojen ylläpitäjien osalta tämä tarkoittaa sitä, että omien toiminnallisuuksien eli moduulien kehittäminen ja lisääminen arkistosovellukseen tulee olemaan entistä helpompaa.

Ulkoasukerros [Manakin](#) kuuluu ohjelmiston perusjakeluun. Manakin on muokattava www-käyttöliittymäkerros, jonka avulla arkistossa olevaa sisältöä esitetään. Ominaisuudessa ei meidän kannaltamme ole mitään uutta, JYX-arkistossa on jo Manakin käytössä. Käytännössä uudistus tuo helpotusta ohjelmiston asennukseen, koska toiminnallisuutta ei enää tarvitse asentaa erikseen. Ulkoasukerrosten osalta uusi jakelu tuo mukanaan sen filosofisen muutoksen, että Dspaceella ei enää ole yhtä oletuskäyttöliittymää vaan sen virkaa ennen toimittanut JSP-UI on nyt vain yksi käyttöliittymä muiden rinnalla.

Uusina rajapintoina esiteltiin myös käyttötarkoituksiltaan osittain rinnakkaiset [SWORD](#) ja [LNI](#) (Lightweight Network Interface), jotka tarjoavat työkaluja DSpace-arkiston sisällön hallintaan. Myös hakutoiminnallisuuden rakennetta on muutettu monipuolisemmin räätälöitävään suuntaan.

Muokattavat julkaisukanavat kuuluvat version 1.5. perusjakeluun. Tähän uudistukseen pätee meidän osaltamme sama asia kuin Manakinin osalta eli olemme käyttäneet muokattavien julkaisukanavien toiminnallisuutta jo jonkin aikaa. Toki toiminnallisuuden sisältyminen perusjakeluun helpottaa tulevia asennuksia.

Julkaisukanavan muokkaamista havainnollistettiin esittelemällä kaksi uutta julkaisuprosessiin liitettävää vaihetta. Toinen niistä oli yksinkertaiset käyttörajoitukset mahdollistava vaihe. Vaiheessa julkaisija voi rajata julkaisun käyttöä joko ajallisesti tai IP-alueen perusteella. Toinen esitelty uusi vaihe oli ”[Pubmed step](#)”, jonka avulla saadaan julkaisun PubMed ID:tä käyttämällä haettua julkaisun tiedot paikalliseen DSpace-arkistoon. Toiminnallisuudella ei ole sellaisenaan JY:n kannalta käyttöä, mutta se antaa hyvän pohjan vastaavanlaisten muita läheteitä käyttävien tiedonhakuvaiheiden suunnitteluun, esim. ISI Web of Science –step. Olinkin jo asian tiimoilta yhteydessä HY:n Terkon DSpace-kehittäjään Mika Stenbergiin, ja sovimme, että kehitämme yhdessä tiedonsiirtotoiminnallisuutta Dspace-pohjaisen arkiston ja Web Of Siencen välille. Toteutuessaan tekniikka mahdollistaa sen, että kansainvälisissä arkistoissa olevat paikallisen yliopiston tutkijoiden julkaisemien artikkelien tiedot voidaan haravoida omaan arkistoon ja sitä kautta muihin järjestelmiin, esim. Tutkaan.

Uuden version julkaisemisen vaikutukset JYX-kehitykselle ja julkaisukanavahankkeelle

JYX:n päivittäminen uuteen versioon tulee ajankohtaiseksi lähitulevaisuudessa. Tulemme toimeen nykyisen version kanssa jonkin aikaa, koska kaikki meille olennaiset toiminnallisuudet on asennettu jo nykyiseen versioon. Välitöntä reagoitua vaatii Julkaisukanavahankkeessa kehitetyn sopimusjärjestelmärajapinnan päivittäminen toimimaan uuden version kanssa. Koska Dspacen versiopäivitys tuo merkittäviä muutoksia Dspacen ja Manakinin lähdekoodiin se tarkoittaa muutoksia myös rajapinnan lähdekoodiin. Ajankohtaiseksi asian tekee se, että sama rajapinta on käytössä JY:n lisäksi Tampereen, Helsingin ja Oulun Dspace-asennuksissa ja meidän täytyy pystyä tarjoamaan uusi versio rajapinnasta siinä vaiheessa kun ko. Yliopistoissa ollaan siirtymässä uuteen Dspace-versioon. Alkuvaiheessa tämä tarkoittaa yliopistojen päivitysaikataulujen kartoittamista ja rajapinnan kehittämisaikataulun suunnittelemista niiden mukaan.

10.4.2008/Pertti Pitkänen

