

ALA:n mies kertoo: raportti ALA:n kesäkokouksesta 2008

29.8.2008

Juha Hakala
juha.hakala (at) helsinki.fi

ALA:n eli American Library Associationin vuosittaiset kesäkonferenssit ovat maailman suurin kirjastoalan tapahtuma. Valtaosa yli 20.000 osallistujasta on amerikkalaisia, ulkomaisia osallistujia on joukossa vain muutamia satoja. Tapahtuma näyttelyineen vetää paikalle myös laajan joukon kustantajia, kirjailijoita ja kirjastoihin liittyvän teknologian toimittajia. Digitaalisen aineiston pitkäaikaissäilytykseen liittyvät kysymykset houkuttelivat minut kokoukseen ensi kertaa. ALA:ssa oli kuitenkin monipuoliset mahdollisuudet tutustua myös keskeisiin kirjastoalan sovelluksiin ja niiden kehitysnäkymiin.

Osallistuin ALA:n kokoukseen Anaheimissa, Kaliforniassa kesä-heinäkuun vaihteessa 2008. ALA-kokoukseksi väkeä oli paikalla niukasti, vain 22 047, joista näytteilleasettajia noin 5000. Vähennystä edellisvuoteen verrattuna oli useita tuhansia, mutta ensikertalaiselle mieleenpainuvinta oli edelleen osallistujien määrä: missään muualla tuskin näkee näin paljon kirjastonhoitajia samalla kertaa.

Kun kolmannes ison maan kirjastoväestä kerääntyy yhteen paikkaan, houkuttelevat he puoleensa muutakin väkeä. Suuret ja vähän pienemmätkin kustantajat näyttäytyivät, ja heidän mukanaan kirjailijat (juhlavieraana Khaled Hosseini), kirjoitustaitoiset näyttelijät (Jamie Lee Curtis), näyttelijät (Diahann Carroll) ja muuten tunnetut henkilöt (Ron Reagan, ei tosin isä vaan poika, joka on amerikkalaisille tv:n keskusteluohjelmista tuttu). Olipa ALA:ssa myös Dean Koontz, jonka pahin kauhukokemus lapsena on konferenssijulkaisun mukaan ollut kirjaston kirjan kadottaminen.

Kaikkineen ALA on melkoinen speaktaakkeli, josta pystyy seuraamaan vain murto-osaa. Minua luonnollisesti kiinnosti eniten tekniikka.

Suomessa ainutlaatuinen pitkäaikaissäilytyksen suunnitelma

ALA-matkani rahoitti Ex Libris, joka pyysi minulta esitystä digitaalisen aineiston pitkäaikaissäilytyksestä käsitelleeseen sessioonsa. Kuulijoissani herätti mielenkiintoa erityisesti se, että Suomessa kaikki muistiorganisaatiot ovat opetusministeriön Kansallisen digitaalisen kirjaston hankkeessa rakentamassa yhteistä pitkäaikaissäilytysjärjestelmää. Näin laajaan yhteistyöhön ei muualla ole vielä päästy. Onkin mielenkiintoista nähdä, miten PAS-toimintaa tullaan muualla organisoimaan. Digitaalisten arkistojen rakentaminen ja ylläpito on haasteellinen tehtävä, jossa yhteistyö on voimaa.

PAS-sessiossa esiteltiin ensimmäistä kertaa julkisesti Ex Libriksen Digital Preservation System –sovellusta. Ohjelmiston betatestaus aloitettiin Uuden Seelannin kansalliskirjastossa kesäkuussa. Tarkoitus on saada järjestelmä tuotantokäyttöön vuoden 2008 aikana.

Valitettavasti demo oli sen verran lyhyt, että ne joille DPS ei ollut jo entuudestaan tuttu, eivät luultavasti saaneet siitä yleiskäsitystä. Sovelluksen peer review group'in jäsenenä minulle oli kuitenkin mielenkiintoista nähdä toteutettuna toimintoja, jotka olivat entuudestaan tuttuja vain paperilla.

PAS-sovellus vastaa monimutkaisuudeltaan hyvinkin integroitua kirjastojärjestelmää. Niissä tulee tyypillisesti olemaan OAIS-mallin mukaiset kuusi toimintakokonaisuutta (aineiston vastaanotto, säilytys ja käyttö; säilytys; tietojen hallinta sekä hallinnointi). Toiminnoissa voidaan erottaa useita erilaisia rooleja

PAS-toiminnan vastuuhenkilöistä tiedon luovuttajiin ja arkistoidun aineiston jakelijoihin.

Jos OAIS-mallikaan ei ole kuulijalle tuttu, siihen perustuvan järjestelmän selittäminen on vaikeaa. Esimerkiksi aineiston vastaanottoon liittyvien toimintojen esittely vastaisi tällöin integroidun kirjastojärjestelmän hankintamodulin esittelyä henkilölle, joka ei tunne kirjastojärjestelmiä lainkaan.

Ex Libriksellä riittää siis haastetta sen selittämisessä, mitä he ovat oikein rakentamassa, ja keille tämä tuote parhaiten soveltuu. Hieman vastaava tilanne heillä oli muutamia vuosia sitten portaalisovelluksen kanssa. Aika näyttää, tuleeko DPS:stä omalla alueellaan yhtä suosittu kuin MetaLibistä ja SFX:stä, vai ryhdytäänkö pitkäaikaissäilytyksessä ylipäätään käyttämään erillissovelluksia. Houkutus ratkaista pitkäaikaissäilytyksen ongelma tavallisella digitaalisen aineiston hallintasovelluksella kuten DSpaceella on suuri, kun kaupan hyllyllä ei ole vielä vaihtoehtoja.

Sähköisen pitkäaikaissäilytyksen järjestelmät keskustelussa

Kokousmatkani PAS-orientaation vuoksi osallistuin myös Digital Preservation Interest Groupin kokoukseen. Ex Libris –session tavoin sekin oli houkutelut väkeä ensisijaisesti suurista tieteellisistä kirjastoista. Asialistalla oli muun muassa TRAC (Trusted repositories Audit and Certification, katso <http://www.crl.edu/PDF/trac.pdf> (<http://www.crl.edu/PDF/trac.pdf>)) ja sen käyttöönotto Yhdysvalloissa ja muualla maailmassa. Lisäksi keskusteltiin termien määrittämisestä – englantia puhuvillekaan ei ole selvää, miten esimerkiksi digital preservation hahmotetaan. Tässä malliksi termin lyhyt ja keskimittainen määritelmä:

Digital preservation combines policies, strategies and actions that ensure access to digital content over time.

Digital preservation combines policies, strategies and actions that ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time.

Molemmissa määritelmissä käytettävyyden varmistaminen on pitkäaikaissäilytyksen tavoite ja onnistumisen kriteeri. Bittien tallentaminen on toki välttämätön, mutta ei vielä riittävä ehto onnistumiselle.

Eurooppalaisena oli mukava kuulla, että amerikkalaisten mielestä vanha manner on PAS-asioissa heitä edellä. Käytännössä tällä viitataan esimerkiksi Ison Britannian Digital Curation Centre –toimintayksikköön (<http://www.dcc.ac.uk/> (<http://www.dcc.ac.uk/>)), EU:n PAS-hankkeisiin sekä Hollannin kansalliskirjaston jo vuosia tuotannossa olleeseen PAS-järjestelmään. Näistä kaikista Preservation Interest Groupin väki tuntui olevan hyvin perillä. Yhdysvaltoihin halutaan DCC:tä vastaava keskus – missäpä sellaisesta ei olisi hyötyä – sekä lisää kokemusta käytännön toiminnasta.

Varsin yleinen PAS-ratkaisu tuntuu Yhdysvalloissa olevan DSpaceen tai Fedoraan perustuva, yhden tai korkeintaan muutaman yliopiston yhteinen järjestelmä. Kysyin käytetäänkö näitä sovelluksia PAS-tarkoituksiin sen vuoksi että ne nyt sattuvat olemaan käytettävissä vai siksi, että ne oikeasti hoitavat homman. Tapaamisen puheenjohtaja totesi, että edellinen tulkinta on hänen mielestään oikea. Digitaalisen aineiston hallintasovellukset yleisesti ottaen sisältävät vain OAIS-mallin aineiston vastaanotto-osiossa määritellyn toiminnallisuuden, tai oikeammin vain osan siitä. Kaikki muu, erityisesti pitkäaikaissäilytykseen liittyvät piirteet, puuttuu näistä ohjelmista - toistaiseksi.

Muita OAIS-toimintoja on rakennettu Fedoraan VTLS:n voimin ja vastaavia DSpace-pohjaisia kehittämishankkeita on tietävästi käynnistymässä. Mutta onko DSpaceen ja muiden vastaavien ohjelmistojen toimintalogiikka sellainen, että se on saumattomasti laajennettavissa PAS-järjestelmäksi? DPS-järjestelmän kehitystyötä seurattuani olen jossakin määrin skeptinen – DSpace ja DPS – ja myös DigiTool ja DPS - poikkeavat toisistaan merkittävästi toimintaperiaatteiltaan ja rakenteeltaan.

Tiedustelin muilta interest groupin kokoukseen osallistujilta miten paljon USA:ssa on käytännön kokemusta arkistoitujen tiedostojen käsittelystä (kopiointi / migraatio). Ilmeisesti varsin vähän, eikä

kukaan ollut vielä konvertoinut todella suuria tiedostomääriä.

Tässäkin suhteessa Eurooppa, käytännössä kollegamme Norjan kansalliskirjastossa jotka konvertoivat parhaillaan miljoonia TIFF-kuvia JPEG2000-formaattiin, edustavat terävintä kärkeä. Toivotaan ettei se osoittaudu bleeding edge'ksi. Mo i Ranan väki teki siellä vieraillessamme selväksi, että menetelmät, jotka toimivat muutamille sadoille tai tuhansille kuville eivät välttämättä enää toimi sadoilla tuhansilla kuvatiedostoilla.

Kirjastojen sovellukset murrosvaiheessa

Yksi ALA:n mielenkiintoisimmista anneista on kokouksen yhteydessä järjestettävä näyttely. Järjestelmätoimittajille se on niin sanotusti pakkopullaa: kaikkien on oltava siellä mukana, sillä poisjäänti aiheuttaisi välittömästi huhuja järjestelmätoimittajan tulevaisuudesta.

Menemättä yksityiskohtiin voi todeta, että integroitu kirjastojärjestelmä on historiaa. Tarjolla olevien isojen ja pienempien kirjastosovellusten määrä kasvaa jatkuvasti. "Isolla sovelluksella" tarkoitan esimerkiksi Voyagerin tai MetaLibin kaltaisia ohjelmia. Pieneksi taas voi lukea vaikkapa VTLS:n ohjelmiston, joka mahdollistaa Virtua-pohjaisen FRBR-tietokannan rakentamisen perinteisen kirjastotietokannan päälle niin, että käyttäjät voivat siirtyä FRBR-tietokannasta perinteiseen näyttöluetteloon ja päinvastoin varsin saumattomasti.

Kirjastoille tarjolla olevien vaihtoehtojen lisääntyminen merkitsee sitä, että ilman kokonaiskuvaa ja kunnollista strategista suunnittelua joudutaan helposti eksyksiin. Tarjolla olevista vaihtoehdoista pitää pystyä valitsemaan ja toteuttamaan ne, jotka ovat palvelujen kehittämisen kannalta tärkeimpiä. Ja koska järjestelmien määrä kasvaa, käyttöönoton yhteydessä on syytä pitää mielessä hankinnan ja ylläpidon kustannukset. Hyvin hajautettu voi tulla myös hyvin kalliiksi.

Kansalliskirjaston tietohallinnon toimenpideohjelmassa on painopisteiksi valittu muun muassa yhteinen tiedonhakupohjainen järjestelmä sekä PAS-sovellus. Tässä ALA-kokouksessa ei tullut vastaan mitään, joka olisi saanut minut epäilemään näiden valintojen relevanssia. Mutta kirjastoautomaatio on tällä erää siinä määrin turbulentissa tilassa, että muutoksia on tarpeen seurata tavallistakin tarkemmin.

Avoin lähdekoodi on tullut kirjastojen järjestelmiin jäädäkseen. Joissakin tapauksissa järjestelmätoimittajat hyötyvät näistä sovelluksista – Lucenea käytetään hakukoneena monissa uusissa hakujärjestelmissä, ja monet ovat tehneet tai tekevät yhteistyötä Index Datan kanssa. Uusin esimerkki on OCLC:n ja Index Datan ALA:ssa julkistettu yhteishanke, jossa WorldCatiin rakennetaan uusia hakuominaisuuksia (<http://www.oclc.org/news/releases/200821.htm> (<http://www.oclc.org/news/releases/200821.htm>)). Mutta avoin lähdekoodi voi olla järjestelmätoimittajille myös kilpailija; konkreettisimmin tämä ilmenee digitaalisen aineiston hallintaohjelmistoissa, joissa DSpace ja Fedora syövät kaupallisten tarjoajien markkinoita.

Avoimen lähdekoodin integroituja kirjastojärjestelmiä on ollut tarjolla jo vuosia (Koha & Evergreen; <http://www.koha.org/> (<http://www.koha.org/>) & <http://open-ils.org/> (<http://open-ils.org/>)), ja niiden käyttäjiä tukevia yrityksiä on perustettu. Se, milloin sovellusten ja tukiohjelmaorganisaation toiminta on niin vakaata, että 350.000 käyttäjän järjestelmään voidaan harkita tämältyypistä ratkaisua, onkin sitten eri asia.

Ex Libris luottaa avoimeen arkkitehtuuriin

Järjestelmätoimittajille sovelluskentän hajoaminen tuntuu olevan vaikea asia hahmottaa muuten kuin lisääntyneinä markkinointimahdollisuuksina. Rajapinnoista puhuu moni; esimerkiksi tätä nykyä suosittu Polaris-järjestelmän rajapintoja kuvataan osoitteessa <http://www.polarislibrary.com/products-services/api.html>, mutta kovin lyhyesti ja pintapuolisesti.

Ex Libris on omaksunut open platform –strategian, jossa avoimien ohjelmistorajapintojen sekä avoimen lähdekoodin merkitys korostuvat. Tavoitteena on, että kaikki yrityksen sovellukset koostuvat jatkossa erillisistä ja itsenäisistä komponenteista. Tältä pohjalta järjestelmien käyttäjät – niin organisaatiot kuin

niin sanotut loppukäyttäjätkin – voivat ainakin periaatteessa yhdistellä saman eri valmistajien sovelluksia sekä omaa kehitystyötä parhaan kokonaistuloksen saavuttamiseksi.

ALA:ssa suurelle yleisölle esitelty Open platform -strategia houkuttelee eniten niitä organisaatioita, joilla on riittävästi omia kehittämisresursseja ja dynaaminen asiakaskunta, joka vaatii uusia palveluita. Muille kelpaa hyvin "musta laatikkokin", kunhan se on vakaa.

Ex Libriksen tuotteissa on toki ollut avoimia rajapintoja 90-luvulta lähtien. Mutta uusi strategia edellyttää, että rajapinnat ovat hyvin dokumentoituja ja johdonmukaisesti toteutettuja kautta koko sovellusvalikoiman. Tähän tavoitteeseen pääseminen edellyttäne vielä runsaasti töitä, varsinkin kun pyrkimyksenä on tarjota sekä Ex Libriksen omia että käyttäjäorganisaatioiden tuottamia sovelluksia.

Open platform –sessiossa esiteltiin myös uutta työkalua, jonka avulla kirjastot voivat julkistaa ja tarjota muiden käyttöön kehittämäänsä apuohjelmia. Esimerkkisovelluksina näytettiin 200 hengen yleisölle Ere Majalan kehittämää välineitä.

Liitäntöihin tarvitaan standardien tukea

Avoimien rajapintojen määrittely on vaikeaa, ellei ole standardeja joiden päälle rakentaa. NISO: on ja muihin standardointijärjestöihin kohdistuu melkoisia odotuksia. Niihin vastaaminen voi olla helppoa silloin kun kaikki työhön osallistujat puhaltavat yhteen hiileen, eikä kilpailevia ehdotuksia ole. Valitettavasti asiat eivät aina ole näin hyvin.

NISO:n SUSHI (Standardized Usage Statistics Harvesting Initiative; <http://www.niso.org/workrooms/sushi>) edustanee onnellisten tähtien alla syntynyttä standardia. Monien muiden uusien NISO-standardien tavoin SUSHI:n käyttöönotto tehostaa back office –töitä. Samaan kategoriaan kuuluvat esimerkiksi asiakastietojen siirtoon soveltuva NCIP (NISO Circulation Interchange Protocol), josta on rakenteilla versio 2 (http://www.niso.org/standards/ballots/ncip_v2_change_summary.pdf).

Standardin kehittäjät toivovat että siitä tulee edeltäjänsä suosittu. Suomessa SUSHIa ja NCIPiä huomattavasti tunnetuimmin tunnettaneen helmikuussa 2008 julkaistu SERU (Best Practices for Shared E-Resource Understanding (<http://www.niso.org/publications/rp/RP-7-2008.pdf>)), joka

"offers publishers and libraries the opportunity to save both the time and the costs associated with a negotiated and signed license agreement by agreeing to operate within a framework of shared understanding and good faith."

NISO on aktiivisesti mukana myös ONIX-kausijulkaisuformaattien kehittämisessä vastuuorganisaation eli EDItEURin kumppanina. Näitä formaatteja ovat (lisätietoja katso <http://www.editeur.org/onixserials.html>):

- SPS (Serials Products and Subscriptions),
- SOH (Serials Online Holdings) sekä
- SRN (Serials Release Notification)

Nämä määrytykset ovat osittain vielä työn alla, eikä ole selvää miten laajalti ne otetaan käyttöön kustantajien ja kirjastojen järjestelmissä. Jo kansallisen ONIX-vastuunsa pohjalta Kansalliskirjasto seuraa tilannetta.

Edellä listatuilla standardeilla ei ole kilpailijoita, eikä juuri kampittajiakaan. Muutamissa standardointihankkeissa näistä ei ole pulaa. Tämä kävi selväksi ALA:n Identifier-sessiossa, jossa keynote speaker John Kunze markkinoi itse kehittämäänsä ARC-tunnistetta ja resoluutiopalvelua, joka on DOI:n ja URN:n kilpailija. Ainakin allekirjoittaneelle jäi epäselväksi miksi ARC olisi parempi kuin muut resoluutiopalvelut, ja voi olla että sama epäily jäi kalvamaan muutaman muunkin kuulijan mieltä. Todennäköistä kuitenkin on, että DOI, URN ja ARC elävät verkossa enemmän tai vähemmän

rauhanomaisesti rinnakkain vielä vuosia.

Lopuksi

En ole aiemmin vierailut ALA:n kesäkokouksessa. Kulttuurishokki ei ollut kovin suuri, koska takana oli sentään yksi ALA:n talvikonferenssi hyisessä Philadelphiassa joitakin vuosia sitten. Eikä konferenssin kokoakaan oikein huomaa, sillä ALA on haarautuvien polkujen puutarha: samanaikaisesti on käynnissä kymmeniä työryhmä- ja muita kokouksia, joissa on paikalla muutamia kymmeniä tai korkeintaan satoja ihmisiä. Sisällöllisesti ALA on hyvin antoisa, kunhan suunnittelee etukäteen mihin menee.

Ainutlaatuisin matkakokemus oli hotelli: koska se oli osa Disneylandia, aamiaista "terrorisoi" joukko Disney-hahmoja. Tilaisuus saada itsestään kuva banaani kourassa ja Hessu toisessa käsipuolella oli varmasti ainutlaatuinen, mutta jätin sen väliin ilman suurempia tunnonvaivoja.

Pyydämme lähettämään palautteen suoraan kirjoittajalle:

juha.hakala(at)helsinki.fi