



Melinda

Tutkittujen järjestelmien teknisiä löydöksiä

Henri Mäkilä
Kansalliskirjasto

Libris XL (1/5)

- Pääperiaate Ruotsissa: tietueet luodaan ja muokataan selainkäyttöisellä kuvailutyökalulla Libris XL:ssä, josta ne replikoituvat paikallisiin kantoihin
- Tietueet ovat Libris XL:n kannassa linkitetyn datan formaatissa (sovellettu Bibframe/RDF + JSON-LD)
- Tietueita ei tallenneta kantaan Marc 21 -muotoisina lainkaan
- Muunnos Marc 21 -> Bibframe -> Marc 21 on häviöllinen
 - kenttänumerot voivat muuttua
 - indikaattoreita, osakenttiä ja välimerkkejä saattaa tippua
 - testeissä kuvailudataan ilmestyi ruotsalaisen kuvailukäytännön piirteitä, koska muunnossääntöjä ei ole tässä vaiheessa lokalisoitu

Libris XL (2/5)

- Toimiva testi-instanssi asennettu KK:n omalle palvelimelle syksyllä 2021
- REST-rajapinta valmiina
 - liikuteltavien tietueiden formaatti JSON-LD
 - tietueet noudettavissa myös Marc 21:ksi konvertoituina
- Haravointi kattavasti tuettu OAI-PMH:n kautta
 - tietueet saa ulos useassa formaatissa (MarcXML, RDFXML, DC, JSON-LD)
- Hakurajapinta (Webbsök) rakennettu avoimen koodin Elasticsearchin päälle
 - Z39.50/SRU:n kautta haut eivät ole täysin reaaliaikaisia, vaan kohdistuvat kannasta ajastetusti otettuun Marc 21 -dumppiin
 - Melinda-integraatioissa SRU-hakuja tehdään runsaasti ja lähes kaikkien tietueoperaatioiden yhteydessä

Libris XL (3/5)

- Tietueiden massatuontia varten erillinen ohjelmistomoduuli
- Tuonteja testattu mm. Fennican ja PIKI-kirjastojen aineistoilla (N ~ 3 000 tietuetta)
 - Marc 21 -muotoiset tietueet voidaan tuoda sellaisenaan
 - tietueet tallentuvat kantaan Bibframe/RDF-konvertoituina
 - nopeus (konversio + tallennus + indeksointi) koeajoissa melko vaatimaton 400 tietuetta/min, mahdollisesti paranee ympäristön optimoinnilla
- Tuotantokäytössä Ruotsissa n. 10M tietueen kanta, suorituskyky Melinda-käytössä olisi varmistettava lisätesteillä
- Osakohteiden käsittely tuonneissa? Melindan erikoisuus: emoilla jopa tuhansia artikkeliosakohteita
- Formaattimuunnosten lokalisointi työlästä
- Tietokannan toteutus minimaalinen: kaikenlaiset tietueet (bibit, holdingsit, auktoriteetit) tallentuvat yhteen relaatiotauluun

Libris XL (4/5)

- Replikointiin on Ruotsissa kolme tapaa:
 - batch export = yleensä yöajo, tietueet ilmestyvät Marc 21 -muodossa KB:n tai paikalliskirjaston palvelimelle (parametrisoitavissa, mitä haetaan)
 - Libris XL:n kannan aikaväliharavointi OAI-PMH:lla
 - tietueiden nouto marc_export-moduulilla
- Melinda-käytössä replikointi hoidettaisiin todennäköisesti jollain em. tavalla tai näiden räätälöidyllä yhdistelmällä

Libris XL (5/5)

- Käyttäjien autentikointi OAuth2:lla
- Kuvailussa kolmitasoinen käyttäjämalli, oikeudet kirjastokohtaisia
 - Admin = pääkäyttäjä, voi luoda uusia käyttäjiä
 - Cataloguer = kuvailija
 - Registrant = oikeus rekisteröidä holdingseja
- Käyttäjämalli siis melkein sama kuin Alephissa nykyään
- REST-rajapintakäytössä kirjastoille myönnetään käyttöoikeusavain (bearer token)
 - kirjasto lähettää avaimen rajapinnalle tunnistautumista edellyttävien kutsujen yhteydessä

Folio (1/4)

- Folion metadatatavaranto on toteutettu moduulilla Source Record Storage (SRS)
 - periaatteessa *formaatti-agnostinen* = ei ota kantaa tietueiden formaattiin
 - käytännön toteutuksissa tietueet tallennetaan kuitenkin toistaiseksi vaihtomuotoisessa Marc 21 -formaatisa
 - myös rajapinnat (mm. REST) olettavat, että kannassa tietueet ovat Marc 21:tä
 - kuvailutyökaluksi integroitava esim. Sinopia
 - linkitetyn datan ominaisuudet saadaan aikaan joko Marc-mappauksilla tai tallentamalla Bibframe-muotoisia tietueita esim. Sinopian omaan rinnakkaiskantaan
 - Folion omassa roadmapissa natiivi linkitetyn datan tuki 2023-25

Folio (2/4)

- REST-rajapintana toimii moduuli QuickMarc (tietueiden luonti, muokkaus ja poisto), liikuteltavien tietueiden formaatti Marc 21
- Haravointia varten OAI-PMH
 - saatavilla olevat formaatit DC ja MarcXML
 - linkitetyn datan haravointia ei toistaiseksi saatavilla
- Hakurajapintana z2Folio (Z39.50-palvelin)
- Laaja käyttöoikeusmalli (Folio Permission Model)
 - moduulien oikeudet säädettävissä hienojakoisesti toiminto- ja käyttäjäkohtaisesti
- Replikointiominaisuudet olisi tehtävä Melinda-käyttöä varten kokonaan omatoimisesti OAI-PMH:n avulla

Folio (3/4)

- Testeissä käytetty ulkoisen toimittajan tekemää asennusta, selvitetty myös mahdollisuutta käyttää KK:n omaa testipalvelinta
- Tietueiden massatuontia testattu koeajoilla (N ~ 8 000 tietuetta) vaihtelevin tuloksin
 - Data Import -tuontimoduulia ei koeasennuksissa saatu toimimaan vakaasti
 - vain osa tietue-eristä meni kantaan nopeasti ja virheettömästi
 - muun kehittäjäyhteisön saamia tuloksia optimoidussa ympäristössä:
 - 50 000 tietueen vienti: 1 h 40 min
 - 50 000 tietueen päivitys: 2 h 32 min
- Testien perusteella ei voitu varmistaa, onko Folion metadatavaranto riittävän stabiili ja suorituskykyinen Melinda-käyttöön

Folio (4/4)

- Järjestelmäarkkitehtuuri vahvasti modulaarinen
 - lähes 60 erillistä moduulia/konttia
 - moduulien välillä Okapi-väliohjelmisto
 - kehittäjäyhteisö kansainvälinen ja aktiivinen mutta myös hajanainen, osaamistasoltaan vaihteleva ja heikosti koordinoitu
 - osa moduuleista selvästi keskeneräisiä ja puutteellisesti dokumentoituja
- Melinda-käyttöön tarvittaisiin lähinnä vain metadatatavaran ja käyttäjähallintaan liittyvien moduulien toiminnot
- Asennettavuus ja ylläpidettävyys ovat testauksen aikana osoittautuneet useista näkökulmista ongelmallisiksi
- Hiomaton ja kyseenalaiseen perusratkaisuun pohjautuva arkkitehtuuri olisi Melindan tuotantokäytössä huomattava riski



Kiitos!

melinda-posti@helsinki.fi

Melinda-wiki: <https://www.kivi.fi/display/melinda>

<https://twitter.com/Melindapalvelu>