ISBD-konversio PIKI-tietokantaan

Konversion pohjana oli Kansalliskirjaston ISBD-konversio. Metiva-wikin sivulla on nähtävillä myös Konversiosuunnitelma ja konversio-ohjelma, jonka avulla pystyy tarkastamaan konversion sääntöjä. Kansalliskirjaston konversiosuunnitelmassa 336- ja 337-kentät muodostetaan periaatteessa neljän kohdan perusteella:

1. nimiön merkkipaikka 06 (tietueen tyyppi)
2. 007 merkkipaikka 00 (aineistoluokka)
3. 008 merkkipaikka 23 (ilmiasu)
4. 245$h (aineiston yleismääre)

245$h poistetaan konversiossa.

Konversiossa muutetaan nimiön merkkipaikkaa 18, jossa kerrotaan luettelointisäännöt a > i (000[18] = i = ISBD).

Kansalliskirjaston konversiosuunnitelmassa ei ole huomioitu osakohteiden 773-kentässä mahdollisesti olevaa emojulkaisun yleismäärettä, ja meillä se oli myös jäädä huomaamatta konversiossa, vaikka se on mainittu Konversiosuunnitelma-sivulla.

PIKI-tietokantaan tehdyt erityistoimet

Usemarconin pohjalta oma konversio-ohjelma

Axiellin ISBD-konversiosta vastasi Marko Happonen. Hänen mielestään vaikutti työläältä ja monimutkaiselta käyttää tässä konversiossa Kansalliskirjaston Usemarconia. Hän teki Usemarcon-taulukon pohjalta ”tietokantaproseduurin” niin, että konversion voi tehdä suoraan kannassa scriptillä ja kannan eräajolla. Proseduuri sallii tarvittaessa myös monimutkaisempaa käsittelyä ja ehdollisuutta, jos on tarpeen.

Marko Happonen nostaa esiin kaksi periaatteellista eroa meidän konversion ja usemarcon –konversion välillä:

1. Usemarcon:in säännöt tärppäävät tietyt ehdot täyttävään sisältöön. Sillä ei ole mahdollista tehdä sääntöä, joka tärppää silloin kun jotkin muut säännöt eivät ole tärpänneet
2. Axiell:in konversio voi ottaa huomioon osakohteiden emonimekkeet

Raporttityökalut

Axiell teki Aurora-ohjelmaan kaksi raporttia: 'Testaa nimekkeen ISBD -konversiota' ja 'Listaus 245h–kentistä.

Raportin 'Testaa nimekkeen ISBD -konversiota' avulla pystyi testaamaan, miten yksittäinen tietue muuttuu konversiossa.

– työkalu osoittautui hyödylliseksi välineeksi muotoiltaessa konversiosääntöjä

Listaus 245h-kentistä osoitti, että tietokannassa oli paljon virheitä 245-kentässä:

- käytetyt yleismääreet olivat virheellisiä (genrejä, esineen nimiä, itse keksittyjä termejä jne.)
    - 245 00 $aAavikkotiellä$h[Runo].
- hakasulkeet puutuivat yleismääreen ympäriltä

- 245-kenttä oli ”rakennettu” virheellisesti:
    - oli jäänyt pois kenttäerotin ja tietoelementin tunnus ($c) ennen vastuullisuusmerkintöjä
    - puuttui tarpeellisia välimerkkejä kuten /-merkki ennen vastuullisuusmerkintöjä tai osakenttämerkintä $b ennen muuta nimekettä
    - välimerkit olivat väärin tai puuttuivat kokonaan

Yleismääreet: 245$h-kenttä

Kaikkiaan 245h–osakenttä löytyi 132 974 tietueesta (ks. liite) ja niistä löytyi paljon virheitä. Pystyimme raporttityökalun avulla korjaamaan virheellisiä kenttiä ja yhdessä Axiellin kanssa laatimaan sääntöjä kenttien koneelliseen korjaamiseen.

Kentän 245 virheet olisivat voineet aiheuttaa paljon virheitä konversioon ja paljon tärkeää tietoa olisi voinut kadota konversiossa (mm. tapaukset, joissa 245$h-kenttään oli inhimillisten virheiden vuoksi mennyt vastuullisuustiedot tai joissa 245$h-kenttään oli laitettu vastoin sääntöjä aineiston lajityyppi) ja siksi piti ensin korjata kentässä olevat virheet. Osa korjauksista tehtiin manuaalisesti, suurin osa korjattiin yhdessä Axiellin kanssa viilattujen korjaussääntöjen avulla:

Esimerkkejä korjauksista, jotka räätälöitiin 245-kenttää varten:

- Jos kentän 245h alussa ei ole välimerkkejä, mutta loppuhakasulkeen perässä on välimerkkejä, jotain tekstiä ja taas välimerkkejä > siirretään hakasulkeiden perässä olevat välimerkit edellisen osakentän (yleensä 245a) loppuun.
- Jos hakasulkeiden perässä on muutakin tekstiä kuin välimerkkejä, tekstistä tehdään 245b -osakenttä.
- Jos 245h:tä seuraa jo ennestään 245b, niin tuo ylimääräinen data tulee olemassa olevan 245b -kentän alkuun.

- Lisätään /$c–merkki yleismääreen jälkeen

    - Jos 245$h-kentässä on jokin ”genrettävä” termi kuten [laulu], [runo], [satu], [leikki], [laululeikki], se säilytetään siten että korvataan merkintä $h merkinnällä $b ja hakasulkeet säilytetään.

Musiikin DVD:t ja Blu-rayt

Yleensä 336/337-kenttiin:

336$aKuva (liikkuva ; kaksiulotteinen).$aMusiikki (esitetty)
337$avideo

Mutta musiikin dokumenttielokuvat:

336$aKuva (liikkuva ; kaksiulotteinen)
337$avideo

- Ehto: jos pääluokkana on
  78.993 (Populaarimusiikin edustajien henkilöhistoria),
  78.991 (Länsimaisen taidemusiikin edustajien henkilöhistoria) tai
  78.992 (Etnomusiikin edustajien henkilöhistoria)

Osakohteilla oli ongelmia: osakohteilta puuttuivat yleismääreet

- dvd/blu-ray-levyjen osakohteilla näyttää muodostuvan toinen osakenttä $aMusiikki (esitetty) niihin osakohteisiin, joissa osakohteessa ei ole yleismäärettä 245h-osakentässä, mutta yleismääre löytyy 773h-osakentästä. Tämä näyttää onnistuvan sekä niiden osakohteiden osalta, joissa on ykl-luokka 084-kentässä, että niiden osalta, joissa ei ole 084-kenttää, vaan se haetaan emosta Mutta nyt siis tutkitaankin 773h –kenttä ja jos sitä ei löydy, niin sitten emon 336 -kenttä. Jos osakohteella ei ole 245h, eikä 773h–osakenttää, ja emolla on 336–kentässä ”Musiikki (esitetty)”, niin osakohteen määreeksi asetetaan ”videolevy”. Sääntöjen tulkinta etenee sen jälkeen normaalisti. Jos osakohteen oma luokka ei ole musiikki, niin tuota ”Musiikki (esitetty)” tietoa ei tule osakohteeseen vaikka se emolla on.
- Jos osakohteella ei ole omaa luokkaa, niin se otetaan emolta.
- Ylimääräisiä hakuja emon 336:een tulee enintään 300000 tietueen osalta.
- Luokan haku emosta on jo ollut tehtynä. Sitä ei tarvitse tehdä kuin silloin kun määreenä on teksti ”videotallenne”. Oikeassa konversiotilanteessa emolta ei löydy enää 245h–kenttää, koska emo on jo konvertoitu.

Moniviestimet

Aineistolajin 'o' käsittely näköjään puuttui Usemarcon-versiosta. Lisättiin PIKI-konversioon.

Pohdittiin mahdollisuutta määritellä moniviestimille tarkempia 336/337-kenttä, mutta tällä aikataululla ei ollut aikaa rakentaa sääntöjä, koska 300-kentän tiedot vaihtelevat niin paljon. Nyt kaikille moniviestimille hyväksyttiin ”Useita sisältötyyppejä” ja ”Useita määrittäviä laitteita”.

Konversio-ohjelmasta puuttuvat yleismääreet

Huomattiin, että kaikki tietueet eivät konvertoituneet, vaikka yleismääre oli oikein. Lisättiin 245$h-kentälle arvoja ja määrittelin niille 336/337-kentät

- Konversio etsii sanoja äänite, nuottijulkaisu, video, heijastuskuva, kartta-aineisto, moniviestin, kooste, mikro, elektroninen ja kuva. Viimeinen sana ”kuva” niin, että se ei voi olla sanan loppu.
- Näyttää siltä, että nämä kaikki termit jäävät käsittelemättä, vaikka ne ovat oikein: Arkistoaineisto, Elokuva, Esine, Hologrammi, Kohokirjoitus, Nuottikäsikirjoitus, Valokuva, Videotallenne
- Laadittiin konversiosäännöt / korjattiin käsin:

  Arkistoaineisto
  - 336__$aTeksti
  - 337__$aei välittävää laitetta
  - Näitä on niin vähän (13), että voin tehdä käsinkin, jos koodaaminen vie aikaa? Ilmoittele!

  Elokuva
  - vain yksi tietue, teen käsin

  Esine
  - 336__$aEsine
  - 337__$aei välittävää laitetta

  Videotallenne: tapaus 1
  - 336__$aKuva (liikkuva ; kaksiulotteinen)
  - 337__$avideo

  Videotallenne: tapaus 2

  Jos tietueelle on 084$a ensimmäisenä toistumana 78.*/lu, niin
  - 336__$aKuva (liikkuva ; kaksiulotteinen). $aMusiikki (esitetty)

        - 337__$avideo

Näille ei tarvitse tehdä mitään

- Hologrammi
    - Ei tietueita
- Nuottikäsikirjoitus
    - Ei tietueita
- Valokuva
    - Ei tietueita

Aineistolajin muutosta ei tehdä yleismääreen perusteella

Kansalliskirjaston säännöissä muutetaan aineistolajeja kentän 245h perusteella ja luodaan sen mukaisia puuttuvia 007 -kenttiä. Axiellin konversiossa jätettiin ko. toiminnot pois, koska silloin pitäisi myös tyhjätä 008 -kentän positiot 18 - 34 ja laittaa aineistosta riippuen ehkä jotain tilalle.

Puutteelliset koodaukset, 336/337-kenttiä ei syntynyt

Kaikille tietueille ei saada puutteellisten koodausten vuoksi syntymään 336/ 337-kenttiä. Jos 336/337–kenttiä ei saada syntymään, niin lisätään ”Määrittelemätön sisältötyyppi” ja ”Määrittelemätön välittävä laite”.

Huomattiin, että jos tietueessa 008[23/24]-kentässä oli ilmiasussa merkki "_" = ei mikään seuraavista > ei syntynyt 336/337-kenttää Aurorassa "_" = alaviiva = ei mikään seuraavista. Vika korjattiin

- Usemarcon:
  008/24/ | 337(new)$a | If (S = ' ' or S = '|' or S = 'd' or S = 'f' or S = 'r') Then 'ei välittävää laitetta'
- PIKI: Käsiteltiin konversiossa merkit ' ', '_' ja '|' synonyymeinä

Osakohteiden 773$h-kenttä

Osakohteisiin ei ole luetteloitu yleismäärettä 245h-osakenttään. Yleismääre on ollut 245-kentässä vain emossa. Mutta 773-kentän h-osakentässä yleismääre on osakohteissa, ainakin yleensä. Voi olla tietueita, joista yleismääre puuttuu 773-kentästäkin osakohteessa.

Määre otettiin 773$h-kentästä, jos osakohteessa ei ole omaa 245$h-kenttää. Emo konvertoituu ennen osakohdetta, joten tehtäessä osakohteen konversiota emossa ei enää ole 245h:ta. Silloin siellä kylläkin on 336.

Tekstiaineiston 007-kenttä

Kenttä 007 tekstiaineistolta olisi pitänyt poistaa, mutta se jätettiin PIKI-konversiossa.

Luettelointitason muuttaminen

Pohdittiin myös luettelointitasokoodin muuttamista nimiössä. Tällä hetkellä vanhoissa tietueissa se saattaa puuttua kokonaan tai olla virheellinen. Nykyään valmiille tietueille PIKIssä laitetaan tasokoodiksi 4, tason 7 olisi voinut laittaa keskeneräisille tietueille. Koska määrittelyyn liittyi ongelmia, tämä jätettiin ISBD-konversion ulkopuolelle.

Tasokoodia muutetaan kuitenkin lennossa poimittaessa tietueet OAI –indeksiin (Finnaan), joten ne menevät jollain tavalla korjattuna Melindaan.

Funktiokoodien avaamien

PIKI-tietokannassa olevien funktiokoodien avaamista ISBD-konversion yhteydessä pohdittiin myös, mutta koska koodien avaamista käytävä keskustelu oli vielä luetteloinnin työryhmissä kesken ja PIKIssä käytetyissä funktiotermeissä oli niin paljon virheitä, niin suunnitelmasta luovuttiin.