

# Konversiossa havaittuja ongelmia

Ohessa joitakin Melindan konversioissa havaittuja ongelmia ja toimenpiteet niiden pohjalta

## Korjatut viat

- Konvertoimattomista asiasanoista, jotka siirrettiin uuteen kenttään ilman sanastokoodia, saattoi syntyä samanlaisia kenttiä (erityisesti 653), jos asiasana esiintyi useamman kerran alkuperäisessä tietueessa. Ongelma korjattu ja myös sanastokoodittomat täysin identtiset kentät poistetaan konversion jälkeen
- 567-kentistä piti konvertoida myös kentät, joissa ei ole sanastokoodia. Korjaus versioon 1.0: kentät konvertoidaan samoilla säännöillä kuin ysa /allärs-sanastokoodin sisältävät kentät
- konversio muutti 655 z-osakentän termit 655 g-osakenttään, jos kentässä ei ollut paikkatermiä vaan muu asiasana. Korjaus versioon 1.0: aihesanat siirretään nyt 650 a-osakenttään
- Python-ohjelmointikielellä käytettävä pymarc-kirjasto hävitti käsittelyssä datan kentistä, jotka sisältävät MARC21-formaatin vastaista sisältöä (ts. kontrollikenttänä, joku muu kuin 00X-merkitty kenttä). pymarc-kirjastoon tehtiin paikallinen paikkaus konversio-ohjelman versiossa 1.01
- Ohjelma jätti ennalleen kentät, joista yhtään asiasanaa ei pystytty konvertoimaan tai joka sisälsi vain poistettavia asiasanoja. Korjaus versiossa 1.02
- Sanat, joille löytyi sanastossa toinen samanlainen mutta kirjainkoolta eroava sana, tulkittiin "exact matcheiksi" esim. ruotsiksi Kemi <> kemi. Mäitä ei konvertoitu vaan jätettiin sääntöjen mukaisesti sellaisenaan ilman sanastokoodia samaan kenttään. Tällaisia sanapareja löytyy vajaa kymmenen kappaletta ja kaikki Allärs-sanastosta. Korjaus versiossa 1.03
- tietueet, joista puuttuu 001-kenttä, jätettiin konvertoimatta. Korjaus versiossa 1.03: 001-kenttään puutteelliset tietueet konvertoidaan
- ohjelma ei liittänyt yhteen sisällöltään identtisiä 382-kenttiä, joissa on toisistaan eroava \$8-osakenttä. Korjaus versiossa 1.04
- SLM-sanastosta on muutamia sanoja, joissa pää- ja ohjaustermit ovat samat, esim. "todellisuuspohjaiset elokuvat". SLM-sanaston kohdalla sanalle annettiin tulkinta päätermin mukaan, vaikka muitakin vaihtoehtoja olisi. Lisäksi samanlaiset ohjaustermit tulkittiin vain yhden termin mukaan. Korjaus versiossa 1.04. Huom! "todellisuuspohjaiset elokuvat" ei ole enää 13.7.2019 jälkeen SLM-sanastossa altLabelina.
- 8-osakentillä linkittämisessä oli vika: jos tietueessa oli asiasanaketjuja ja niiden jälkeisissä kentissä yksittäisiä asiasanoja, jälkimmäiset asiasanat saivat saman \$8-osakentän kuin viimeinen asiasanaketju. Korjaus versiossa 1.05
- Sanastoissa on altLabeleita, joille löytyy identtinen prefLabel. Näissä tapauksissa aiemmissa versioissa altLabel tai prefLabel katsottiin joko monitulkintaiseksi tai konvertoitava termi tulkittiin prefLabelin mukaan. Version 1.05 korjauksen jälkeen kaikki tällaiset termit konvertoidaan prefLabelin mukaiseen YSO-vastineeseen. Huomattiin kuitenkin ristiriitaisuus version 1.04 ja 1.05 välillä: SLM-sanastossa toimittiin eri tavalla kuin Musa/Cilla- sekä Ysa/Allärs-sanastoissa. Versioon 1.06 käytännöt yhdenmukaistettiin ja kaikki termit, jotka löytyvät sekä alt- että prefLabelina, viedään tarkistuslistalle ja jätetään konvertoimatta.
- musiikki- ja elokuva-aineiston 650 \$y-osakentissä olevat paikkatermit siirtyivät luomis aikaan viittaavaan 388-kenttään. Versiossa 1.07 650 \$y-osakentistä löytyvät paikkatermit siirretään 257-kenttiin (elokuvat) ja 370-kenttiin (musiikki), jos kyseessä on luomispaikka

## Korjauksia ja puutteita, jotka vielä tekemättä

### Tekniset puutteet

- MARCXML-tiedostojen luku pysähtyy, jos XML-rakenne on viallinen. Mahdollinen korjaus: tietueiden luku yksi kerrallaan

### Ei-toivotut piirteet konversiossa

- ohjelma ei liitä yhteen sisällöltään identtisiä konvertoituja kenttiä, joissa on sekä \$8- että \$9-osakenttiä (tämä on jätetty korjaamatta, koska voi olla riski, että kenttien yhdistämisestä syntyisi vääriä tulkintoja). Nyt tuloksena saattaa olla joitakin turhia rivejä.

### Puutteet sanastoissa

- Samanniminen paikka puuttuu yso-paikat ontologiasta, jolloin termi konvertoituu väärään paikkaan.
  - Tähän ei ole korjausta. Listaamme alla havaittuja tapauksia:
  - Viborg (Tanskassa) Viborg / Viipuri, Karjalan Kannaksella
  - Lima (Ruotsissa) Lima (Perussa)
- Ruotsinkieliset soitinten nimet katoavat tietueesta
  - soitimet, jotka siirrettiin soitinkookooppaano-kenttään 382
  - Jos ruotsinkielinen termi löytyi seko-sanastosta, tietueeseen kirjoitettiin seko-sanaston käytettävä termi
  - Koska seko-sanasto on yksikielinen, ruotsinkielistä vastinetta ei luotu kenttään 382, mutta konvertoitu osakenttä poistettiin
  - Tähän ei ole korjausta tällä hetkellä.
    - Asia voidaan korjata jälkikäteen muuttamalla seko-sanasto kaksikieliseksi