Tietokantojen siivoustarpeet ennen KOHA asennuksia

Tähän dokumenttiin kootaan listaa tietokantojen siivoustarpeista ennen KOHA asennuksia

Asiakastiedot

- Asiakkaat, joilla on useita viivakoodeja
  - Asiakkaalla voi olla samassa asiakasryhmässä useita viivakoodeja, joista vain yksi on aktiivinen
  - Asiakkaalla voi olla useita aktiivisia viivakoodeja, jotka ovat eri asiakasryhmissä
    - Pitää tarkistaa, otetaanko mukaan KOHAan
- Asiakkaat, joilta puuttuu viivakoodi tai viivakoodi(t) on passivoitu
  - KOHAssa asiakasta haetaan ensisijaisesti viivakoodilla
- Asiakkaat, joilta puuttuu PIN-koodi
  - JYssä näitä päivitetään tietyin väliajoin automaattisesti, mutta koko ajan tuntuu syntyvän myös uusia
- Asiakkaiden HETUt kannattaa selvitellä ja siivota ennen migraatiota
  - Kyselyllä esiin virheelliset HETU-muodot
  - Mitä tehdä erilaisten HETU-muotojen kanssa, kannattaako siivota syntymäaikaan? (tämä vielä mietinnässä)
    - Kannattaa tarkistaa, että esim. vaihto-opiskelijoita ei turhaan roiku asiakkaina, esim. jos eivät lainanneet 3 vuoteen (tai mikä rajoitus kussakin kirjastossa on). Esim. kyselyt tilastointiryhmän / asiakasryhmän perusteella.
  - HETUttomien asiakkaiden käsittely
    - Hetulaan voi viedä vain suomalaisen muotoisia HETUja. Muille asiakkaille pitää miettiä käytäntö. Ehdotus: HETU ei pakollinen tieto, asiakkaan syntymäaika tallennetaan ja lisäksi jonkin sovittuun lisätietokenttään esim. passin nro tai opiskelijanumero. (ks. Pääkäyttäjäryhmän viikkokokous 20180906)
- Asiakkaat, joilla on maksuja (huom! eri käytäntöjä kirjastoissa maksujen perimisestä)
- huomautuskentät
  - migraatiossa Voikkarin eri huomautuskentät (general, address, phone, barcode) mappautuu samaan huomautuskenttään. Voikkarista tulleet eri tyyppiset huomautukset erottuvat alkuliitteellä, jossa kerrotaan, mistä Voikkarin huomautuskentästä tieto on peräisin.
  - KOHAssa käytettävissä mm. 2 osoitekenttää, joihin voi ohjata address -kentän huomautuksen
  - mitä vähemmän tarpeettomia huomautuskenttiä siirtyy, sen helppolukuisempi tuo yhteinen huomautuskenttä tulee olemaan
- Tietosuoja-asetukseen liittyvää asiakastietojen poistamista??
- Kannattaa miettiä, mitkä asiakastietueet viedään KOHAan

Kokoelmatiedot

- Maa- ja kielikoodit
  - nvolk: mitäs noista halutaankaan kaivella?
  - Onko vaikutusta Koha-siirrossa, jos bibin maa- tai kielikoodi puuttuu? Kannattaisiko puuttuvat koodit korjata kuntoon Voikkarissa nyt vai myöhemmin Kohassa? Johanna Miettunen 11.9.18
  - Entä jos omassa Voyagerissa on paljon puuttuvia maa- tai kielikoodeja, miten kannattaa korjata? Voiko tehdä massa-konversioajon /replikointiajon Melindassa tms? Johanna Miettunen 11.9.18 HUOM!! TÄTÄ MUUTEN PITÄÄ POHTIA,ELI TEHDÄÄNKÖ BIB-DATAAN LIITTYVÄT KORJAUKSET JOKAISEN OMASSA KANNASSA VAI VOISIKO NIITÄ TOSIAAN AJAA MELINDAN KAUTTA VAIKKA KAIKKIIN KONVERTOITAVIIN KANTOIHIN YHTÄAIKAA. ONKO MELINDASSA YLIPÄÄNSÄ TYÖKALUJA TÄLLÄISEEN JA MITEN TÄÄ KÄYTÄNNÖSSÄ TOTEUTETTAISIIN?
    - nvolk: mun näppituntuma on, ettei Melindasta kannata tehdä massa-ajoja. Replikointijonot häiritsee yksittäisten luetteloijien työtä ym. Yksittäiset tietuiden korjaukset Melindassa kyllä tippuu nätisti. Mulle on kyllä epäselvää, että onko maa- tai kielikoodien puuttuminen mitenkään kriittistä migraation kannalta. Meillä KK:ssa tosin muut ihmiset (Ari, Olli-Antti ja Joonas) enempi noita murehtii, ja minä vain sumplin Voyagereja, ml. voyager-only-massamuutokset.
- Kahtena merkkinä tallennetut ääkköset
  - bib-tietueissa, mutta myös varastotietueissa?
  - `select distinct bib_id from bib_data where REGEXP_LIKE (dump(record_segment, 1016), '^.* (41|61|4f|6f),cc,(88|8a).*$');`
  - Voiko migraatiossa korjata vai korjataanko jo Voyagerissa? Tietueita on tuhansia...
  - nvolk: voin tehdä skriptin (tai oikeastaan minulla on se jo, se pitää vain paketoida käytettävämpään muotoon), joka muuttaa composed ja decomposed -merkit Voyageissa käytetyiksi (ä=1 merkki, á = a + ´ = 2 merkkiä). Kunkin kannan systeemihoitajan pitää sitten tehdä bulkimportilla päivitysajo. Holdarit ovat vaikeampia päivittää. Skriptillä on helppo generoida korjatutut holdarit, mutta niiden vieminen on vaikeampaa. Periaatteessa interleaved-tiedoston käyttö bulkimportin kanssa toimii. Käytännössä en uskaltaisi käyttää sitä, vaan veisin päivitetyt tietueet sisään batchcat-rajapinnan kautta, mutta se menee jo vaikeammaksi. Itemeillä ei ole selkeää tietuetta, toisin kuin bibeillä ja varastotietueilla, joten niiden ääkkösten korjaaminen pitää tehdä joko käsin tai suoraan kantaan.
- Yhteensidotut nimekkeet, listaamista ja miettimistä, miten toimii KOHAssa
  - listaus
    - listaus holdareista, joilla on useampi bib:
      - ```
        select mfhd_id
        from (
        select mfhd_id, count(*) as c
        from bib_mfhd
        group by mfhd_id
        )
        where c > 1
        order by c desc, mfhd_id
        ```
- Bib-tietueet ilman varastotietueita (tarkistettava onko ongelma KOHAssa)
  - select bib_id from (select bt.bib_id, bm.mfhd_id from bib_text bt left join bib_mfhd bm on bt.bib_id=bm.bib_id) where mfhd_id is NULL order by bib_id
  - Jos varastotietue puuttuu, ei tiedetä missä kokoelmassa teos on
  - Varastokirjastossa on paljon bib-tietueita, joilla ei ole varastotietuetta, mutta niillä on low tag
  - JYKDOKiin ajettu isot määrät vanhempia vapaakappaleita Fennicasta ilman varastotietueita 160818 HM
        ```

- Varastotietueita ilman niteitä
  - select mfhd_id from (select mm.mfhd_id, mi.item_id from mfhd_master mm left join mfhd_item mi on mm.mfhd_id=mi.mfhd_id) where item_id is NULL order by mfhd_id
- Missing, lost library applied
  - select i_s.item_id, ist.item_status_desc, i_s.item_status_date from item_status i_s, item_status_type ist where (ist. item_status_desc='Missing' or ist.item_status_desc='Lost--Library Applied') and ist.item_status_type=i_s.item_status order by i_s.item_id
- Kokoelman ja item typen väärät yhdistelmät
- Itemien viivakoodituplat
  - Näille on olemassa KK:ssa tehty skripti, käyttökomento Voyager-kirjastopalvelimilla:
    `/m1/shared/bin/viivakooditarkistin.perl` *xxxdb*
    jossa xxxdb on tietokannan nimi. Skripti tosin taisi tulostaa väärinä tuplina yhteennidottujen teosten viivakoodit...
- Niteiden tilat ja miten ne siirtyvät KOHAan tarkistettava
- Perussiivouksia ja inventointia kannattaa tehdä
- Julkaisurekisterin tietueiden siivoaminen
- Kokoelmien läpikäyminen ja sellaisten kokoelmien poistaminen, joita ei tarvita
- Suppressoidut kokoelmat, joita on jäänyt roikkumaan (voiko KOHAssa piilottaa tiettyjä aineistoja niin, että ne näkyvät vain virkailijalla - mitä Finnassa näytetään vaikuttaa tähän)
  - Kannattaa miettiä poimintaa suunniteltaessa, saadaanko jätettyä tiettyjä kokoelmia ottamatta mukaan poimintaan
- Kannattaa käydä omaa tietokantaa läpi ja merkitä ylös sellaisia tila- tai lisätietoja, joita on merkitty niteisiin ja holdingseihin.
  - Esimerkiksi huonokuntoisuus on eri paikkoihin merkitty
  - eri tyyppistä huomautustietoa on myös hankintamoduulin sisällä. Tsekatkaa, onko siellä säilyttämisen arvoista. Huomautukset ovat omissa tauluissaan Voikkarissa, eli sisällön saa helposti tsekkaukseen esim. exceliin
- Niteiden ja asiakkaiden tilastointiryhmät - ovatko tarpeellisia ja pitääkö siirtää KOHAan
  - Asiakkaiden tilastointiryhmiähän me tarvittaneen jatkossakin yhteistilaston puitteissa (tästä erillinen sivu kts Kansallisen yhteistilaston tilastoryhmät). Toisaalta nehän varmaan ovat meillä jotakuinkin identtiset ainakin ylätasolla. Niteiden tilastoryhmät (jos siis kyse on Item Statistical Categoies -arvoista) eivät ainakaan Jykylässä ole tilastoryhmiä, vaan tätä toimintoa on käytetty niteiden siirtoihin liittyvänä toiminnallisuutena/huomautuksena (Avohyllystä varastoon, Avohyllyyn, Korvaa avohyllyn kirjan ym.). HM 27.6.18
  - Tritoniassa Item Statistical kategorioilla on merkitty ammattikorkeakoulujen hankkimat niteet - tieto tarvittaneen järjestelmävaihdon jälkeenkin. / ChN 27.6.2018