



Osakohteet aineistonsiirroissa

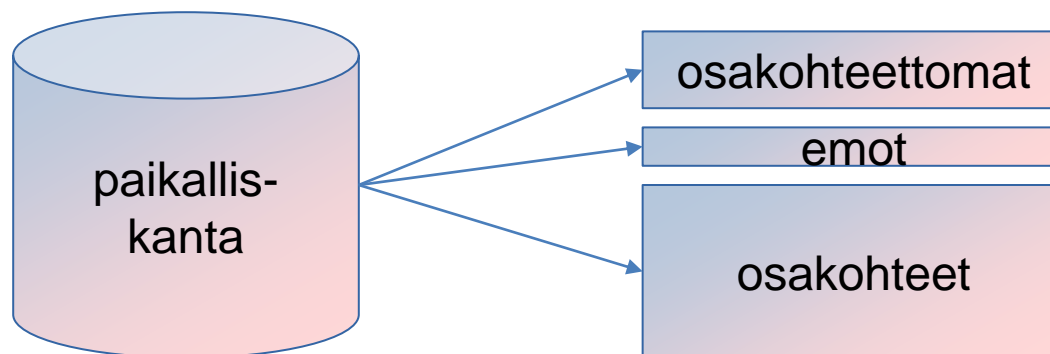
Henri Mäkilä, Melinda-seminaari 3.12.2020
Kansalliskirjasto, Kirjastoverkkopalvelut

Melindan osakohteiden lähteet

- Iso räätäli -mallin aineistonsiirroissa osakohteita on ollut 60–70 % kirjastojen kaikista tietueista.
- Osakohteet enimmäkseen musiikkia eli levyjen raitoja ja nuottien osia. Mukana myös artikkeleita, novelleja, satuja yms.
- Yleisten kirjastojen 1. pilotilta (PIKI) tuli Melindaan suuri määrä uusia osakohteita v. 2015. Toista pilottia (Anders) varten oli nopeasti kehitettävä menetelmä, jolla sieltä tuleva osakohdeaineisto saadaan yhdistymään Melindaan.
- Muita merkittäviä osakohdelähteitä: Arto (siirretty Melindaan 2018), Viola (2019), Yleisradion Fono-aineiston tuonnit.
- Nykyisin Melindassa yli 8 miljoonaa osakohdetta.

Haravointi (Aurora-tietokannat)

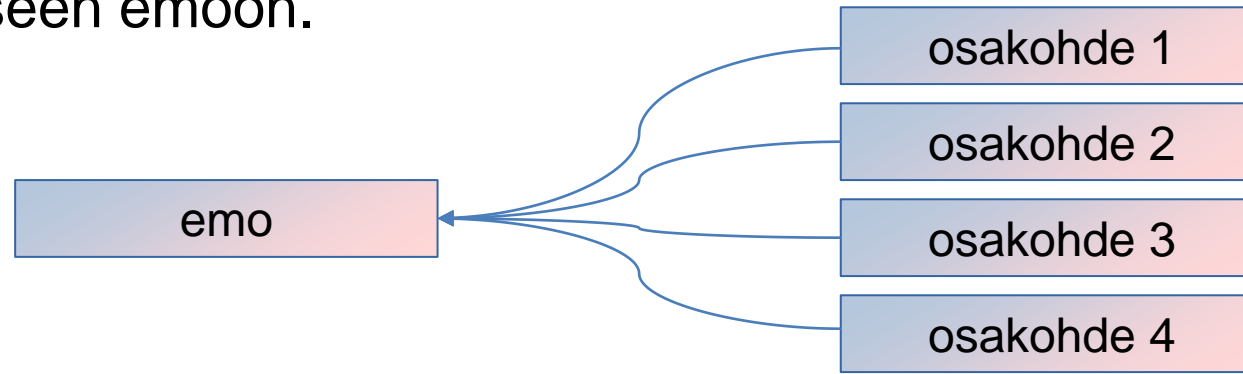
- Paikalliskanta haravoidaan OAI-PMH-rajapinnalla ja aineistosta erotellaan osakohteettomat tietueet, osakohteet ja niiden emot.



- Osakohteiden 773w-linkit tarkistetaan ns. orpojen poikasten varalta. Aurora-kannoissa näitä ei yleensä esiinny.
- Vastinpareja ei voi osakohteille etsiä Melindasta suoraan, koska ne päätyisivät satunnaisten emojen alle.
- Osakohteet on ensin kiinnitettävä oikeisiin emoihinsa.

Osakohdeperheiden muodostaminen

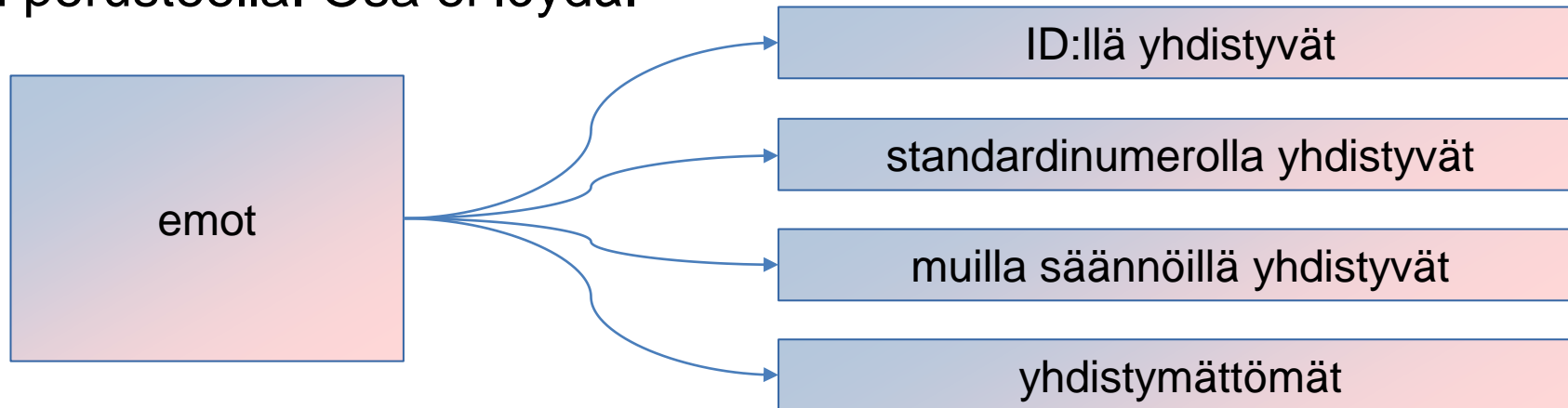
- Perheet kasataan luomalla tietorakenteita, joissa kaikki tietyn emon osakohteet on linkitetty kyseiseen emoon.



- Toisinaan tietorakenteesta tulee varsin suuri. Emolla voi olla satoja ja joskus tuhansia osakohteita.
- Yhdistämistä varten tietorakenteita tarvitaan yhtä monta kuin osakohteilla on 773w:llä linkitettyjä emoja, usein siis kymmeniä tuhansia.

Emojen vastinparien haku

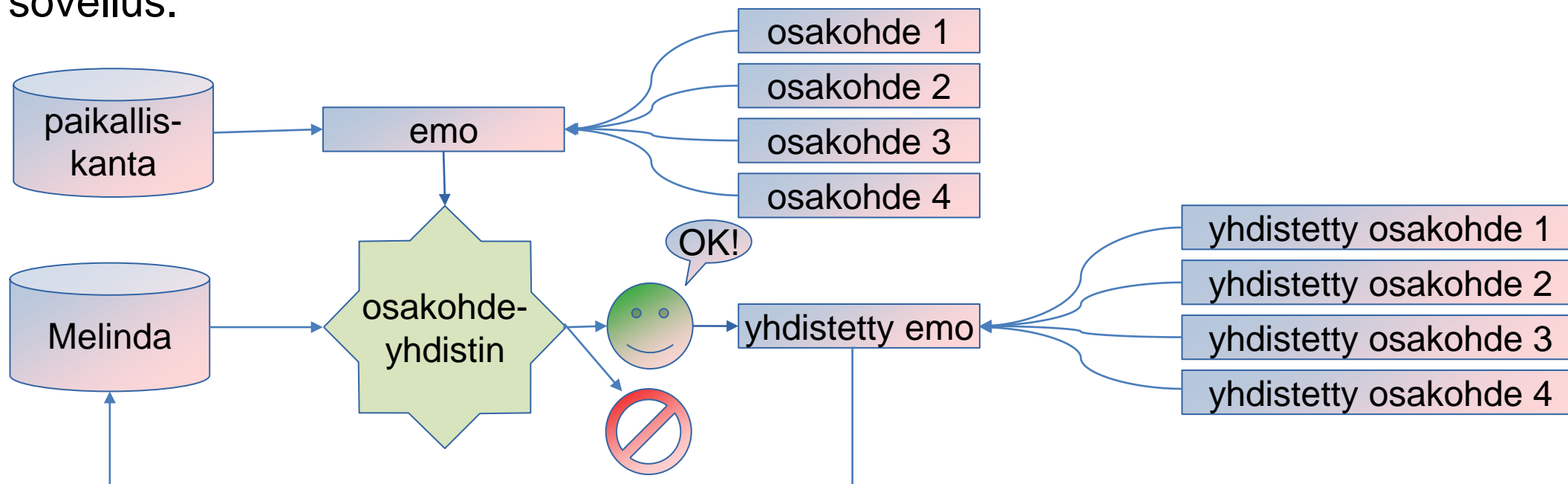
- Perheiden siirto aloitetaan etsimällä emoille Melinda-vastinparit. Suuri osa emoista löytää vastineensa standardinumeron, 035:ssä olevan Melinda-ID:n ja muiden sääntöjen perusteella. Osa ei löydä.



- Emon vastinparin löytyminen ei vielä takaa, että koko emon osakohdeperhe saadaan yhdistymään Melinda-emon perheeseen.
- Jos vastinemo löytyi, kokeillaan seuraavaksi, löytävätkö myös kaikki emon osakohteet oikeat vastineensa Melinda-emon alta.

Osakohdeyhdistin 1/2

- Kokonaisen perhetietorakenteen vastinparin hakua varten on koodattu erityinen sovellus.



- Osakohdeyhdistin vertaa paikallisen emon osakohteita Melindan vastineisiin ja päättelee, voidaanko koko perhe yhdistää. Yhdistetty perhe päivitetään Melindaan.

Osakohdeyhdistin 2/2

- Nykyisessä osakohdeyhdistimessä on rajoitteensa:
 - Automaatio vähäistä: yhdistin osaa käytännössä vain vertailla kahta tietorakennetta. Kaikki muu on tehtävä erillisillä apuohjelmilla tai käsin.
 - Perhe ei yhdisty, jos emon osakohteiden määrä on paikalliskannassa eri kuin Melindassa.
 - Paikallisten osakohteiden pitää olla koodattu osakohteiksi (000/07 = a, b tai d). Osakohde ei saa yhdistyä ei-osakohteeseen.
 - Vertailussa käytettävä pisteytys saattaa hylätä yhdistymisen esim. välimerkkierojen tai kirjoitusvirheiden takia.
 - Hakee tietueita yksitellen Alephin web-rajapinnalta, joten kyselyjä ja tietoliikennettä tulee paljon. Aikaa menee, jos vertailtavia perheitä on runsaasti.
 - Moneen vastinpariin yhdistyvät emot ("multimatchit") ja perheet ongelmallisia.

Ongelmatilanteita 1/2

- Osakohdeyhdistelyn aikana syntyy erityyppisiä tilanteita:
 - **Ideaalitapaus:** Koko perhe (emo + osakohteet) löytää onnistuneesti yksiselitteiset vastinparit Melindasta, yhdistin hyväksyy perheen yhdistämisen.
 - Emo on jo Melindassa mutta ilman osakohteita. Tällöin Melinda-emolle voidaan ajaa paikalliskannan osakohteet suoraan uusina tietueina.
 - Emolla on jo osakohteet myös Melindassa, mutta osakohdeyhdistin hylkää perheen yhdistämisen. Perhe pitää siis ajaa uutena (ensin emo, ja kun sille annettu Melinda-ID on tiedossa, osakohteet sen alle) → saattaa syntyä isokin tuplaperhe.
 - Emo löytää useita vastinpareja (tuplia). Tässä tapauksessa pitää syöttää jokainen perhevariaatio yhdistimelle erikseen ja selvittää, mikä tai mitkä perheet pystyvät yhdistymään ja valita niistä yksi.

Ongelmatilanteita 2/2

- Lopputulos: osakohdeaineisto hajoaa lukuisiin pieniin eriin sen mukaan, miten vastinparien haku ja osakohdeyhdistely on kussakin tapauksessa onnistunut → kirjanpito työlästä, käsin tehtäviä hitaita vaiheita paljon.
- Osakohdeyhdistin toimii nykyisin noin 70-prosenttisesti → jos 10 000 emoa yhdistyy ensin oikein, näistä n. 3 000 perhettä on silti ajettava uusina. Tuplia tulee.
- **Violan erikoiskäsittely:** ekspressiokuvailu (moniemoiset osakohteet) purettiin Melinda-siirrossa: ekspressio-osakohteista kloonattiin yksittäisiä manifestaatio-osakohteita, joilla on vain yksi 773-kenttä.
 - Uusia tavallisia osakohteita syntyi yhtä monta kuin ekspressio-osakohteella on emoja.
 - Alkuperäiset ekspressio-osakohteet on säilytetty Melindassa tietueina, joilla on useita 973-kenttiä, mutta ei yhtään 773:a.

Osakohteiden tuonti jatkossa

- Melindan osakohdeyhdistimen jatkokehitys on lähinnä resurssikysymys. Kehittämistä kaipaavia ominaisuuksia ovat mm.
 - yhdistymistuloksen parantaminen → vähemmän tuplia
 - automaation edistäminen → vähemmän manuaalista työtä
 - koodin optimointi → nopeampi suoritus aika
- Tulevat aineistot siirtyvät Melindaan enenevästi erätuontijärjestelmällä.
 - Toistaiseksi tuodaan vain uusia tietueita.
 - Osakohdetuki erätuontiin? Ei vielä päätöstä.



Kiitos!

melinda-posti@helsinki.fi

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

<https://twitter.com/Melindapalvelu>