



Paikallistietokannan laadun parantaminen Melindan metatiedolla

Ville Huhtala
Helsingin yliopiston kirjasto
29.11.2022

Tekijöiden auktorisointi 100-kentässä

- Tulevaisuuden haasteita metadatalle
 - Linkitetty data tulossa, tarvitaan auktorisoidut toimijat
 - (toivottavasti) ensi vuonna 2023 teosauktoriteetit → tekijä-nimeke auktoriteeteissa vaaditaan, että tekijä on auktorisoitu
- Kirjastoilla ei ole ollut pääsyä toimija-auktoriteetteihin Fennicaa lukuun ottamatta
- Melindaan auktoriteetit tuotiin vuonna 2013
- Suurin osa korkeakoulukirjastoista siirtyi Alma-kirjastojärjestelmään 2020, samalla saatiin käyttöön Asteri
- Helkassa suurin osa kotimaisista toimijoista on auktorisoimatta. **Korjaus vaatisi tuhansien ja tuhansien toimijoiden takautuvaa auktorisointia...**

Melinda

- Tarvittavat auktorisoinnit on mahdollisesti jo tehty Melindan metadatatassa (todennäköisesti Fennican toimesta)
- Näitä tietueita voi myös hakea Melindasta
 - rajoitus, vaatii jonkinlaisen tunnistautumisen, minun tapauksessani yliopiston IP-alueen
- Melindan hakurajapintojen dokumentaatiota:
 - [Fennica/Arto/Viola/Melinda Z39.50 and SRU API](#)
 - [Hakurajapinnat / Search & harvesting APIs](#)

Kirjastojärjestelmä - Alma

- Kiinnostavista tietueista luotiin ensin Almassa setti käyttäen indikaatiosääntöä
 - Vain tekstuaalinen aineisto: nimiö/06-07 = am
 - Julkaisumaa Suomi: 008/15-16 = fi
 - Oltava Melinda-yhteys: 035\$a = (FI-MELINDA)
 - On tekijä: kenttä 100 on olemassa
 - Jo auktorisoituja ei kannata tarkistaa: 100\$0 ei ole olemassa
 - Väitöskirjaa alemmat opinnäytteet pois: 920 (Melindassa 509) kenttää ei olemassa (Helkassa noin 114000)

```
rule "100 ei Asteri, julkaistu Suomessa"
when
  ((existsControl "LDR.{6,2}.am") AND (existsControl "008.{15,2}.fi") AND (exists "035.a.*FI-MELINDA*")
  AND (exists "100") AND (not exists "100.0.*FI-ASTERI-N*") AND (not exists "920"))
then
  set indication."true"
end
```

Käsiteltävät tietueet

- Setissä 205,648 tietuetta, mikä liian paljon tähän työprosessiin
- Tästä perussetistä luotiin pienempiä settejä, jottei käsiteltävää tulisi liikaa - kerralla mielellään <10000 tietuetta käsittelyssä
- Esimerkiksi yhtenä settinä käsiteltiin
 - 1800-luku
 - 1960-luku
 - vuodet 2014-2015 jne.
- **Datasta löytyi myös tupla-035 FI-MELINDA -viitauksia, jotka korjattava**

Helka

LDR 01541cam#a22003374i#4500

001 9925178723506253

005 20220326194615.0

008 981013s1897#####fi#####c#####000#0dfin#c

015 ## \$a f419510\$2skl

035 ## \$a **(FI-MELINDA)000893567**

035 ## \$a (FI-HELKA)2517872-helkadb

041 1# \$a fin \$h swe

100 1# \$a Schybergson, M. G.

Työkalu - MarcEdit

- HULibissa on on jo useita vuosia käytetty MARC-tietueiden käsittelyssä ohjelmaa nimeltä [MarcEdit](#)
 - MarcEdit on tietueprosessoinnin ja käsittelyn "sveitsiläinen linkkuveitsi", josta löytyy monia ominaisuuksia
 - **Ei vaadi ohjelmointiosaamista**
- Tässä tapauksessa oleellinen työkalu oli MarcEditin sisäänrakennettu SRU-klientti
- MarcEditillä voi myös mergetä tietueita, tässä oleellista se, että
 - Helkan tietueessa on Melinda ID 035 kentässä
 - Melindan tietueilla on Melinda ID
- → mergeäminen tapahtuu näitä ID:tä vertaamalla, eli data voidaan yhdistää, kun ID:t mätsäävät

Database Properties

SRU Database Properties:

Name:

Melinda SRU (IP-rajoitettu)

URL:

https://sru.api.melinda.kansalliskirjasto.fi/bibprv

Version:

1.2

Metadata Schema:

marcxml

Search Profile:

Custom



Custom Profile Settings

Title Index:

dc.title

ISBN Index:

bath.isbn

Keyword Index:

melinda.mel

Author Index:

dc.author

ISSN Index:

bath.issn

Subject Index:

dc.subjet

Record # Index:

rec.id

Melindan indeksit:

- rec.id
 - hakee indeksistä ID:llä
- melinda.melindaid
 - edelleenohjaava indeksi

Melindan hakuindeksi, normaalitapaus rec.id & melinda.melindaid

```
LDR 03391cam a2200829 i 4500
001 000893567
003 FI-MELINDA
005 20220927035107.0
008 981013s1897\\fi\\| | | c | | | | | | | | | | \\ | | fin | \\
015 ## $a f419510 $2 skl
035 ## $a (FI-MELINDA)000893567
040 ## $a FI-NL $b fin
041 1# $a fin $h swe
042 ## $a finb
100 1# $a Schybergson, M. G., $d 1851-1925. $0 (FI-ASTERI-N)000083744
```

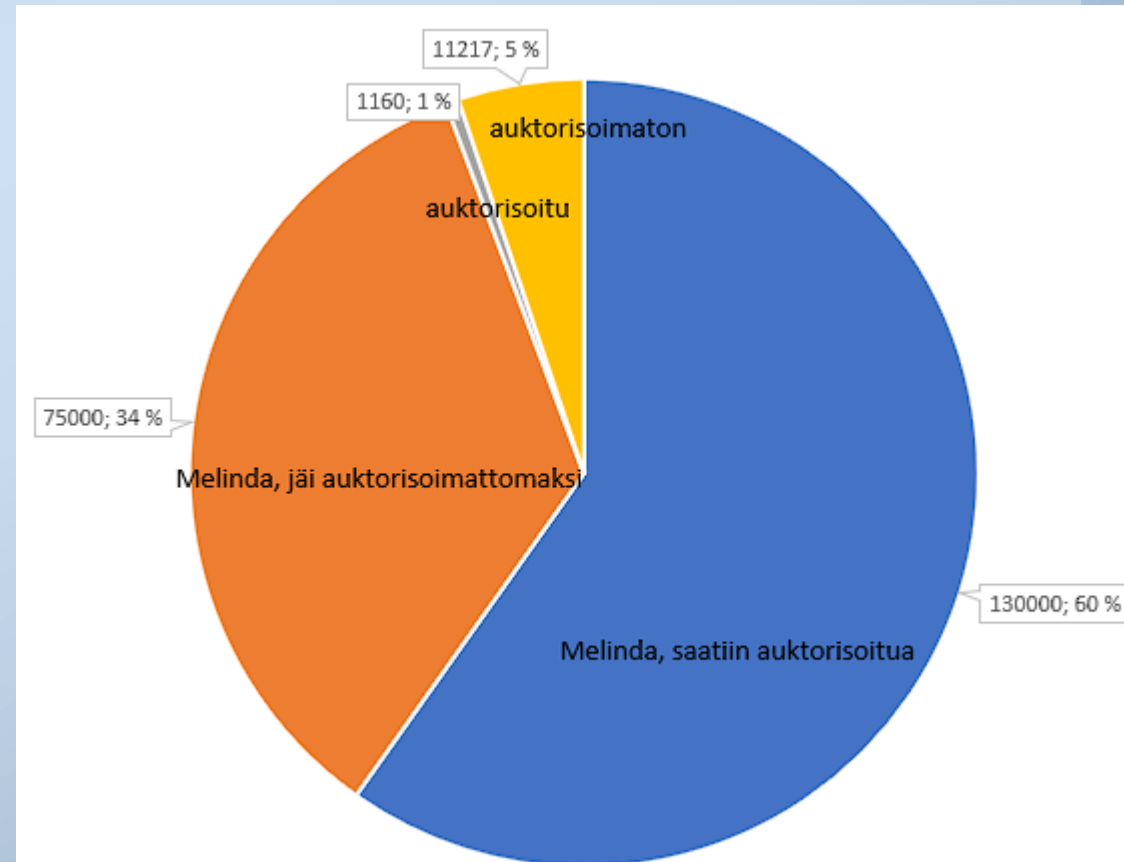

Helkan tietue ↔ Melindan tietue

	A	B	C	D
1	035\$a	HELKA	VRT	MELINDA
2	(FI-MELINDA)000476055	aHentola, Yrjö	TOSI	aHentola, Yrjö
3	(FI-MELINDA)000476117	aJoenpelto, Eeva	EPÄTOSI	aHelle, Eeva
4	(FI-MELINDA)000476121	aHelle, Jarl	TOSI	aHelle, Jarl
5	(FI-MELINDA)000477673	aTarkiainen, V	EPÄTOSI	aTarkiainen, Viljo

- Keltaisella merkityt: tutkittava vielä sen jälkeen, kun tietue importoitu takaisin Helkaan. Tässä tapauksessa menee oikein:
 - 100 1# \$a Joenpelto, Eeva. → 100 1# \$a Helle, Eeva, \$d 1921-2004. \$0 <http://urn.fi/URN:NBN:fi:au:finaf:000079282>
- Mergetään (yhdistetään) tietueet MarcEditissä siten, että
 - Helkan tietue, jossa 035 FI-MELINDA(<ID>) vastaa Melindan tietueen ID:tä
 - Melindan tietueesta otetaan 100 kenttä ja ylikirjoitetaan sillä Helkan tietueen 100 kenttä
- Mergetty tietue importoidaan takaisin Helkaan

Lopputulos Helkassa

- Alussa tietueita, joilla 035MELINDA yhteys ja 100 kenttä auktorisoimaton oli noin 205000
 - **130000 tietuetta saatiin auktorisoitua**
 - 75000 jäi auktorisoimattomaksi
 - koko proseduuri vei noin kaksi viikkoa
- Ilman 035MELINDA yhteyttä 12377 tietuetta
 - joista auktorisoituja 1160
 - auktorisoimattomia 11217



Manuaalisesti korjattuja väitöskirjoja

Vuosi	Kaikki	Auktorisoimattomat	Auktorisoitu	%auktorisoitu
2000	991	241 → 141	750 → 850	75,7 → 85,8
2005	1039	210 → 77	829 → 963	79,8 → 92,7
2010	1000	344 → 155	656 → 845	65,6 → 84,5
2015	777	111 → 90	666 → 687	85,7 → 88,4
2020	698	94 → 41	604 → 657	86,5 → 94,1
YHT	4505	1000 → 504	3505 → 4002	77,8 → 88,8