

AMK-kirjastojen Melindaan tuoma metadatakartunta - loppuraportti

Sisällys

1. Johdanto	1
2. Selvitys.....	1
2.1 Kysely.....	1
2.2 Tekninen toteutus	2
3. Tulokset	3
3.1 Yleisiä lukuja	3
3.2 Uuden aineiston koostumus.....	4
3.2.1 Sisältö- ja mediatyypit.....	4
3.2.2 Kielijakauma.....	5
3.2.3 Julkaisuvuodet.....	6
3.2.4 Asiasanat.....	7
3.2.5 UDK- ja YKL-luokitukset.....	8
3.3 Pois jätetyt aineistot.....	10
4. Yhteenvedo	10

1. Johdanto

Opetus- ja kulttuuriministeriön alaisten ammattikorkeakoulujen kirjastot liittyivät Melindan tiedontuottajiksi syksyn 2012 ja alkuvuoden 2014 välisenä aikana. Ammattikorkeakoulukirjastojen liittyminen muodostaa ensimmäisen vaiheen Melindan kehittämisessä tieteellisten kirjastojen yhteisluettelosta kohti kaikkien kirjastosektoreiden yhteistä metatietovarantoa Kansalliskirjaston koordinoiman ja opetus- ja kulttuuriministeriön rahoittaman *Kansallinen metatietovaranto* -hankkeen puitteissa.

Ammattikorkeakoulukirjastojen kokoelmia on päätyntä Melindaan myös jo ennen metatietovarantohankkeen käynnistämistä yhteiskirjastojen tietokantojen mukana. Tällaisia ovat Kemi-Tornion AMK ja Rovaniemen AMK (Lapin korkeakoulukirjasto yhdessä Lapin yliopiston kanssa), Saimaan AMK (Lappeenrannan tiedekirjasto yhdessä Lappeenrannan teknillisen yliopiston kanssa), Vaasan AMK ja osa Yrkeshögskolan Novian ja Centria ammattikorkeakoulun kokoelmia (Tritonia). Yhteiskirjastojen aineistot eivät ole mukana tässä katsauksessa.

2. Selvitys

2.1 Kysely

Koska metatietovarantoon tulee meneillään olevien AMK-kirjastojen liittymisprosessien myötä uusi kirjastosektori lähes kokonaisuudessaan, hankkeen johdossa heräsi kiinnostus tarkastella liit-

17.12.2014

Kirjastoverkkopalvelut

tyvien organisaatioiden Melindaan tuoma metatietokartuntaa sekä määrällisten tunnuslukujen että sisällön näkökulmista. Selvityksen tavoitteena oli ennen kaikkea tarkastella, millaista aineistoa metatietovarantoon on ammattikorkeakoulukirjastojen myötä tullut uutena. Toinen tavoite selvityksessä oli tarkastella laajemmin ammattikorkeakoulukirjastojen metadatta, sen volyymeja ja uutena tulevan aineiston osuuksia sekä millaisia aineistoja kirjastot ovat jättäneet metatietovarannon ulkopuolelle.

Selvitys toteutettiin kahdessa osassa. Ammattikorkeakoulujen Melindaan tuoma metatietokartuntaa selvitettiin ensimmäisen kerran kesällä 2013, jolloin tuossa vaiheessa liittyneiden kahdenoista kirjaston tilanteesta laadittiin väliraportti¹. Vuoden 2014 lopussa OKM:n hallinnonalaan kuuluvien ammattikorkeakoulujen kirjastojen liittymiset Melindaan on saatu päätökseen, joten Melindaan tullutta metatietokartuntaa voidaan tässä vaiheessa tarkastella kokonaisuutena. Selvitys noudattaa pääpiirteissään väliraportin rakennetta, mutta on itsenäinen dokumenttinsa.

Selvityksen tiedot kerättiin Melindaan tuotua dataa tutkimalla sekä kaikille liittyneille AMK-kirjastoille osoitetulla sähköpostikyselyllä. Yhdeksätoista Melindaan liittyneestä ammattikorkeakoulukirjastosta kolmetoista vastasi kyselyyn. Vastausprosentiksi muodostui näin 68 %.

Kysely sisälsi seuraavat kysymykset:

1. *Millaisia ainealueita kirjastonne kokoelmat kattavat?*
2. *Mitkä ovat kirjastonne painopiste- tai erityisainealueita?*
3. *Millaista aineistoa (aibepiirit ja aineistotyypit) kirjastonne loadissa on tullut uutena Melindaan?*
4. *Millaisista syistä olette jättäneet aineistoa pois Melindasta?*

Kirjastojen vastaukset täydentävät datasta tehtyjen kyselyjen muodostamaa kuvaa, valottavat yleisen tason linjauksia sekä kertovat, millaisia aineistoja metatietovarannon ulkopuolelle on jätetty.

2.2 Tekninen toteutus

Selvityksen tiedonhakuosuus eteni lyhyesti kuvattuna seuraavasti. Jokaisen Melindaan liittyneen ammattikorkeakoulukirjaston tietueet tulostettiin Melindasta erillisiin tiedostoihin Aleph Sequential -formaattissa. Poiminta suoritettiin 12.11.2014 ja se tehtiin tietokantatunnisteen (LOW-tag) perusteella. Poikkeuksen tästä muodostivat Metropolian Arsca-tietokannassa olevat aineistot, jotka eroteltiin Arsca-tietokannan muista tietueista Melindaan vietäessä muodostetun ajotunnuksen avulla. Tämän jälkeen poimitut tietueet sisältäviä tiedostoja käsiteltiin, yhdisteltiin ja analysoitiin ohjelmallisesti tietokannan ulkopuolella. Käsitteily tehtiin Perl- ja Python-kielillä kirjoitettujen skriptien ja Unix-tekstityökalujen avulla. Näin muodostettujen raporttien loppukäsittely ja visualisointi suoritettiin taulukkolaskentaohjelmassa.

¹ Raportti löytyy Metiva-wikistä: <https://wiki.helsinki.fi/display/Metiva/Kansallinen+metatietovaranto+-hanke>

3. Tulokset

3.1 Yleisiä lukuja

Vuosina 2012-2014 liittyneiden ammattikorkeakoulukirjastojen mukana Melindaan tuotiin yhteensä noin 1,2 miljoonaa tietuetta. Tämä luku on kaiken ammattikorkeakoulukirjastoista Melindaan ladatun aineiston yhteismäärä, sisältäen samojen tietueiden toistumat eri tietokannoissa. Erillisiä tietueita tässä aineistossa on yhteensä 706000, josta 65 % (460000) on tullut sisään uusina ja 35 % on yhdistynyt Melindan vanhoihin tietueisiin.

Selite	Luku
Kirjastoja yhteensä	19
Tietokantoja	20
Ladattuja tietueita yhteensä	1,2 miljoonaa
Erillisiä tietueita	706000
Uusina tulleita tietueita	460000

Taulukko 1. Aineistonsiirtojen yleisiä tunnuslukuja

Uutena tulleen aineiston suurta osuutta selittää osittain se, että tähän lukuun on sisällytetty ammattikorkeakoulujen opinnäytetyöt, Ebrary-tietueet sekä osakohdetietueet. Ammattikorkeakoulujen opinnäytetöitä Melindaan on tuotu 51000 (11 % uutena tulleesta aineistosta). Vain osa ammattikorkeakouluista on tuonut opinnäytteensä Melindaan, sillä ne löytyvät keskitetysti myös julkaisurkistopalvelu Theseuksen kautta.

Ebrary-tietueita Melindaan on tuotu 86000 nimekettä (19 % uutena tulleista). Pääosa kirjastoista on jättänyt Ebrary-tietueet vain paikalliskantojensa kautta käytettäväksi. Lisäksi jo Melindaan tuodut Ebrary-tietueet on sittemmin päätetty poistaa metatietovarannosta toistaiseksi kuvailutietojen käyttöoikeuksiin liittyvien epäselvyyksien johdosta.

Uutena Melindaan tulleista tietueista 15 % (68000) on osakohteita, käytännössä äänitteisiin ja nuottijulkaisuihin liittyviä yksittäisiä musiikkikappaleita kuvaavia tietueita. Tämä luku ei kuitenkaan ole kovin käyttökelpoinen, sillä ammattikorkeakoulujen liittymisprosessista erillisenä työvaiheena toteutettava osakohdeaineistojen lataaminen oli selvityksen laatimishetkellä vielä käynnissä.

Edellä mainituista varauksista johtuen opinnäytteet, Ebrary-tietueet ja osakohteet on karsittu tarkemmin tutkittavan aineiston ulkopuolelle. Mikäli nämä aineistotyytit jätetään huomioimatta, ammattikorkeakoulukirjaston uutena Melindaan tuoman aineiston kokonaismääräksi muodostuu 251000 tietuetta. Seuraavien osioiden sisällöllinen tarkastelu rajataan tähän tietuejoukkoon.

Taulukkoon 2 on koottu yhteenveto Melindaan vietyjen AMK-kirjastojen tietokantojen määrällisistä tunnusluvuista. Ladattujen tietokantojen keskimääräinen koko oli hieman vajaat 62000 tietuetta, ilman opinnäyte-, poikas- ja Ebrary-tietueita runsaat 49000. Taulukon viimeisen rivin lukemat ilmaisevat, kuinka suuri osuus tästä lukemasta on ”uniikkietueita”, joissa on vain yksi tietokantatunnus.

17.12.2014

Kirjastoverkkopalvelut

Tieto	Min.	Ka.	Maks.
Tietueita yhteensä	18369	61558	201511
Ebrary-tietueet (%)	0 %	2 %	43 %
Poikastietueet (%)	0 %	4 %	32 %
Opinnäytteet (%)	1 %	7 %	24 %
Muut	17418	49107	85493
Muut (%)	42 %	87 %	99 %
Uniikkeja muista (%)	7 %	23 %	44 %

Taulukko 2. Tietokantakohtaisia lukemia

3.2 Uuden aineiston koostumus

3.2.1 Sisältö- ja mediatyypit

Taulukko 3 kuvaa aineiston yleisimpiä sisältö- ja mediatyypikirjauksia. Uutena Melindaan tulleesta aineistosta valtaosa on painettua tekstiaineistoa. Sähköisten aineistojen ja nuottijulkaisujen osuudet ovat myös merkittäviä, kun taas muut aineistotyyppit ovat määrällisesti marginaalisempia. Videotallenteiden osalta ammattikorkeakoulujen mukanaan tuoma metatietokartunta lienee suhteellisesti merkittävin, sillä Melindassa on ennen AMK-kirjastoja ollut lähinnä satunnaisten opetusvideoiden kuvailutietoja.

Sisältötyyppi : mediatyyppi	Osuus
Teksti : ei välittävää laitetta	64,2 %
Teksti : elektroninen	13,8 %
Musiikki (notatoitu) : ei välittävää laitetta	13,2 %
Kuva (liikkuva ; kaksikulotteinen) : video	2,9 %
Musiikki (esitetty) : audio	2,7 %
Puhe : audio	0,3 %
Kuva (kartografinen) : ei välittävää laitetta	0,2 %
Kuva (liikkuva ; kaksikulotteinen) : heijastettava	0,2 %
Määrittelemätön sisältötyyppi : elektroninen	0,2 %
Useita sisältötyyppejä : useita mediatyyppejä	0,1 %
Useita sisältötyyppejä : elektroninen	0,1 %
Musiikki (esitetty) : video	0,1 %

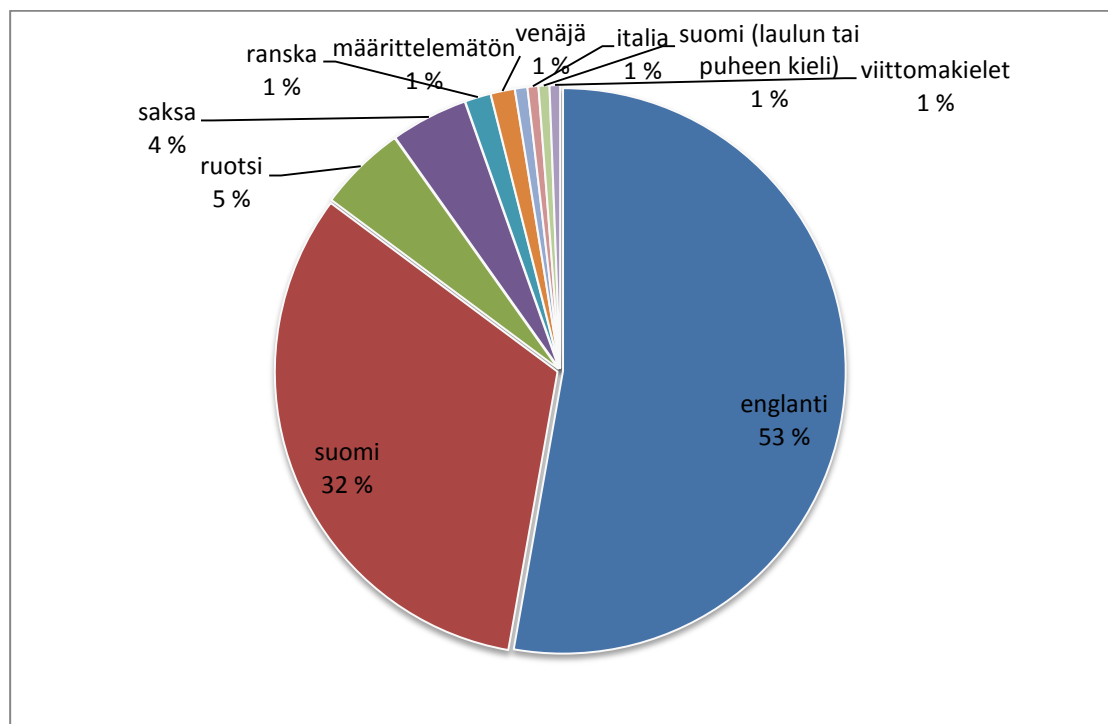
Taulukko 3. Uuden aineiston sisältö- ja mediatyyppejä

17.12.2014

Kirjastoverkkopalvelut

3.2.2 Kielijakauma

Kuva 1 esittää uutena tulleen aineiston kielijakauman piirakkakaaviona. Kuvaan on otettu kymmenen yleisintä kielikirjausta (041-kenttä) siten, että tiivistelmiin ja alkuteosten kieliin viittaavat kirjaukset on jätetty huomioimatta.

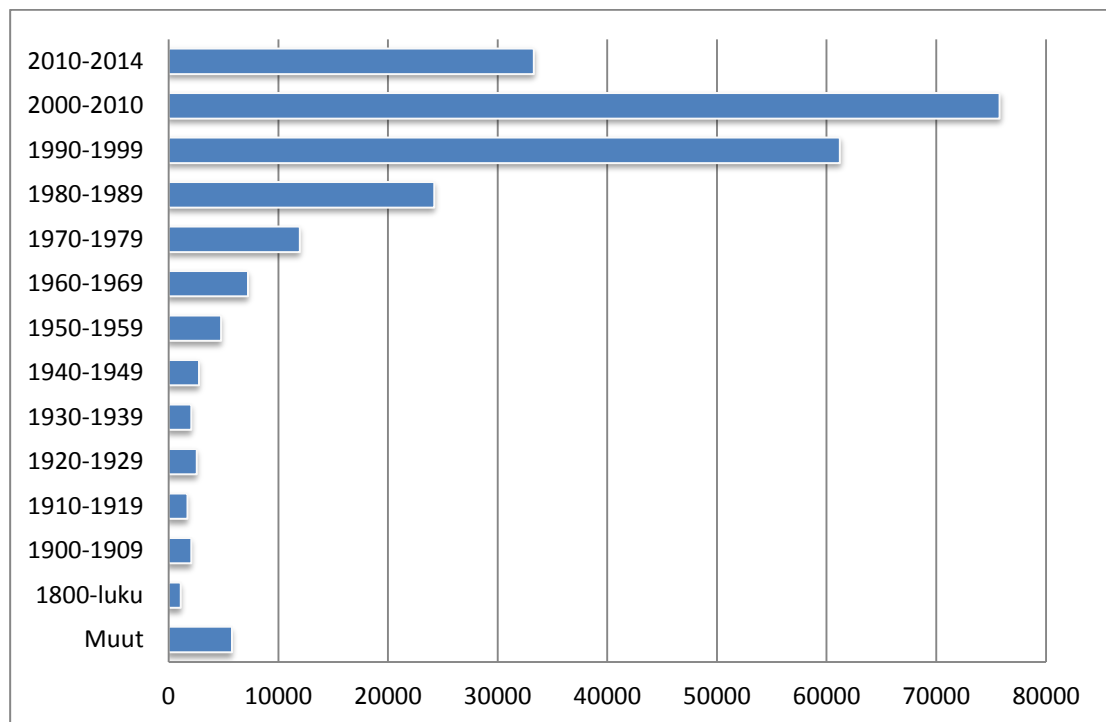


Kuva 1. Uuden aineiston kielet

Valtaosa Melindaan uutena tulleesta aineistosta on englanninkielistä. Suomeksi julkaistua aineistoa on kolmannes, kun taas muiden kielten osuudet jäävät muutamiin prosentteihin kokonaisuudesta. Ruotsinkielisen aineiston osuus on välivaiheen aineistokartoitukseen nähden yli kaksinkertaistunut, mitä selittää ainakin ruotsinkielisen Arcada-ammattikorkeakoulun liittyminen tammikuussa 2014. Musiikkikokoelmien lataaminen näkyy tässä kaaviossa laulu- tai puheäänitteen kieleksi (osakenttä 041\$d) kirjatun suomen nousuna yleisimpien kielikirjausten joukossa yhdeksänneksi. Mielinkiintoista on myös viittomakielten asema kymmenenneksi yleisimpänä kielenä. Ennen ammattikorkeakoulujen mukaan tuloa ei viittomakielistä aineistoa ollut Melindassa käytännössä lainkaan.

3.2.3 Julkaisuvuodet

Kuva 2 esittää uutena tulleen aineiston julkaisuvuosien jakautumista vuosikymmenittäin. Uutena Melindaan tullut aineisto on yleisesti ottaen tuoretta. Lähes puolet (45 %) uusien tietueiden kuvaamista nimekkeistä on julkaistu 2000-luvun puolella. Kymmeniä vuosia vanhan aineiston osuudet ovat pieniä. Luokka ”Muut” kattaa tietueet, joiden julkaisuvuosi ei ole ollut tiedossa tai joissa se on kirjattu puutteellisesti.



Kuva 2. Julkaisuvuodet

3.2.5 UDK- ja YKL-luokitukset

Uutena Melindaan tulleen aineiston sisällönkuvailua tarkasteltiin lopuksi vielä luokituskirjausten kautta. Tarkasteluun otettiin ensimmäisenä kirja-aineiston UDK-luokituskirjaukset, joiden määriä aineistossa kartoitettiin ja peilattiin Kansallisbibliografian UDK-luokituskaavioon³. Taulukkoon 4 on listattu kaksikymmentä yleisintä uutena tulleessa aineistossa esiintynyttä UDK-luokitusta. Yleisimpien kirjausten listaa hallitsevat tekniikkaan, talouteen, pedagogiikkaan sekä erilaisiin terveysaloihin liittyvät luokitukset.

Kirjauksia	Luokka	Selite
4847	004	Tietotekniikka. Tietojenkäsittely. Atk
3999	658	Liiketalous
2238	621	Konetekniikka. Konepajatekniikka. Mekaaninen teknologia. Työstö
2183	65.01	Organisaatiotutkimus. Organisaatioteoria. Johtaminen
1650	621.3	Sähkötekniikka. Voimat. Elektroniikka
1487	37	Kasvatus. Opetus. Koulutus
1404	658.8	Markkinointi. Myynti. Vähittäiskauppa. Tukkukauppa
1358	629	Kulkuneuvot. Liikennevälineet
1226	616	Patologia. Tautioppi
1087	006	Standardit. Normit. Standardisoiminen
990	159.9	Psykologia
943	69	Talonrakennus
938	624	Rakennustekniikka yleensä. Rakennusstatikka. Geotekniikka. Maarakennus. Perustukset. Sillanrakennus. Kalliorakennus. Tunnelinrakennus
867	621.39	Viestintäteknika. Tietoliikennetekniikka
864	615.8	Hoitomenetelmät. Terapia. Psykoterapia. Psykoanalyysi. Kuntoutus. Fysioterapia. Vaihtoehtoiset hoitomenetelmät. Kansanlääkintä
859	339	Kauppa
773	659	Mainonta. Tiedotustoiminta. Joukkoviestintä. Suhdetoiminta.
763	364	Sosiaaliturva. Sosiaalihuolto. Sosiaalityö. Sosiaalipalvelut. Sosiaalivastukset. Kansanterveys. Sairaalat
759	802.0	Englannin kieli

Taulukko 4. Uuden aineiston yleisimmät UDK-luokitukset

³ http://www.kansalliskirjasto.fi/kirjastoala/fennica/fennica_udkkaavio.html

17.12.2014

Kirjastoverkkopalvelut

Uutena tulleen aineiston yleisimpien YKL-kirjauksien listaus taulukossa 5 täydentää UDK-luokitusten antamaa kuvaa. YKL-kirjauksissa musiikkiaineistoihin liittyvät luokitukset ovat selvästi yleisin ryhmä, mutta UDK-kirjausten listallakin vahvasti esillä olevat terveydenhuolto, liiketalous ja tietotekniikka ovat myös edustettuina. Listojen erot kertovat luokitusjärjestelmien erilaisista soveltamisaloista.

Kirjauksia	Luokka	Selite
4956	78.61	Piano
2949	78.322	Yksinlaulu pianon tai muun sointusoittimen säestyksellä
2094	84.2	Suomenkielinen kertomakirjallisuus
1357	78.71	Viulu
1087	69.1	Liiketalouden hallinto
1005	59.21	Hoitotiede. Hoitotyö. Hoitohenkilökunta
720	78.73	Sello
666	78.54	Orkesterit (yleensä enemmän kuin 25 soittajaa)
664	78.3411	Sekakuorit
623	59.2	Yleinen terveydenhoito
592	78.351	Oopperat
572	78.65	Kitarat
562	78.8922	Jazz
551	78.541	Sinfoniat
538	69.3	Markkinointi. Kauppa
501	78.62	Urut
490	61.3	Ohjelmointi. Ohjelmointikielet
448	78.5132	Neljä jousisoitinta
437	78.8911	Rock
433	69.11	Johtamisen tekniikka ja menetelmät

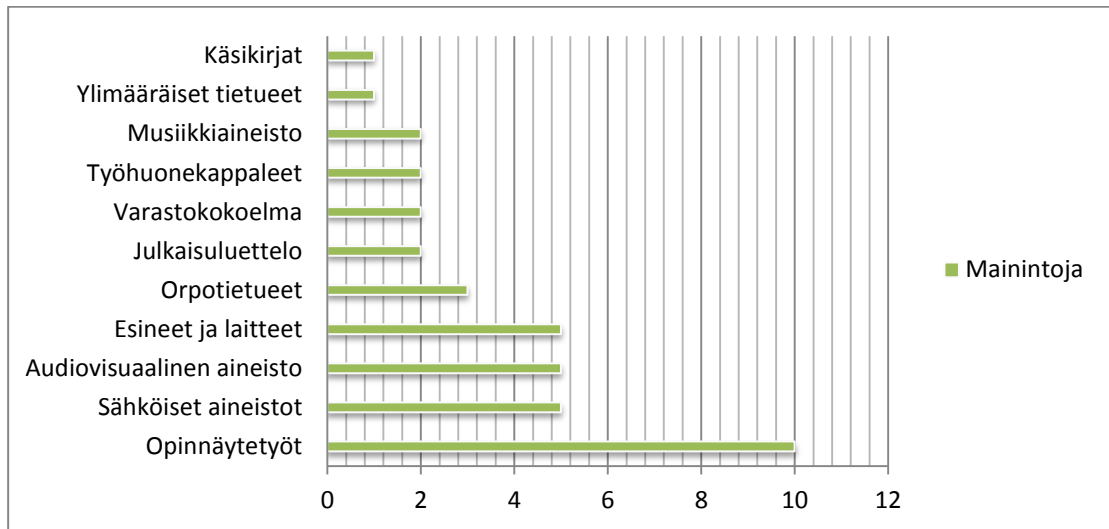
Taulukko 5. Uuden aineiston yleisimmät YKL-luokitukset

Kirjastojen edustajien kyselyssä esittämät näkemykset uutena Melindaan tulleiden tietueiden kattamista aihealueista ovat varsin hyvin linjassa hakukyselyjen tuottamien tulosten kanssa. Kuten odottaa saattaa, kaikki yksittäisten kirjastojen erityisalueet eivät kuitenkaan tule uutena tulleen datan yleisimpien teemojen kartoituksessa esiin. Liiketalous, eri tekniikan alat sekä terveydenhuolto- ja hoitoalat ovat erittäin yleisiä ammattikorkeakoulujen koulutusaloja, joten näistä teemoista Melindaan on karttunut aineistonsiirroissa paljon uusia tietueita. Kirjastojen antamien tietojen mukaan näiden teemoihin kuuluvaa uutta aineistoa on tullut esimerkiksi laivanrakennukseen, apuvälinetekniikkaan, jalkaterapiaan, naprapatiaan, osteopatiaan sekä toimintaterapiaan liittyen. Lisäksi mainintoja uusina Melindaan tulleista aihepiireistä saivat merenkulku, konservointi, rakennuskonservointi, musiikkitekniikka, varhaisiän musiikkikasvatus, maanmittaus, vaatetus, hotelli- ja ravintola-ala, kauneudenhoito, viittomakielentulkkaus, rikosseuraamusala sekä turvallisuusala.

17.12.2014

Kirjastoverkkopalvelut

3.3 Pois jätetyt aineistot



Kuva 4. Tuomatta jätetyt aineistot

Kirjastoille lähetetyn kyselyn vastausten perusteella opinnäytteet ovat selvästi yleisin Melindaan tuomatta jätetty aineistoryhmä. Syyksi tähän ilmoitettiin ammattikorkeakoulujen Theseus-julkaisuarkiston ensisijaisuus opinnäytteiden kuvailuun ja varastointiin, minkä vuoksi niitä ei ole kirjastoissa aina viime vuosina luetteloitu kirjastotietokantaan enää ollenkaan. Monet kirjastot mainitsivat myös erilaiset käyttörajoitukset syyksi aineistojen jättämiselle yhteistietokannan ulkopuolelle. Osalla kirjastoja lainausrajoitukset koskevat painettuja opinnäytetöitä, mutta erityisesti sähköisiä ja audiovisuaalisia aineistoja, kirjojen työhuonekappaleita ja käsikirjastoja. Melindan yleisten linjausten mukaisesti myös paikalliskäyttöiset esineet, kuten lukulaitteet ja lautapelit on jätetty ottamatta Melindaan. Mainituiksi tulivat lisäksi ”orpotietueet”, joilla tässä yhteydessä tarkoitetaan tietueita, joihin ei liity paikalliskannassa varastotietueita tai jotka on piilotettu käyttäjiltä. Lisäksi mainintoja saivat oman organisaation julkaisuluettelot sekä varastokokoelmien puutteellisesti kuvailtu ja järjestetty aineisto. Myös Melindan näkökulmasta ylimääräiset tietueet, kuten kausijulkaisujen yksittäisille numeroille paikalliskantaan tehdyt bib-tietueet saivat kyselyssä yhden maininnan. Selvityksen ensimmäisessä vaiheessa (kesällä 2013) kyselyyn vastanneet kirjastot mainitsivat tässä yhteydessä myös osakohteita sisältävät musiikkiaineistot, jotka teknisistä syistä tuotiin Melindaan erikseen pääaineistonsiirtojen jälkeen. Koska osakohdekoelmien tilanne on sittemmin muuttunut, näitä mainintoja ei huomioitu enää tässä yhteenvedossa.

4. Yhteenveto

Ammattikorkeakoulujen liittyminen Melindan tiedontuottajiksi on laajentanut metatietovarannon katetta ja monipuolistanut sen sisältöä. Metatietokartunnassa painottuvat ammattikorkeakoulujen yleisimmät painopistealueet kuten tekniikan ja liiketalouden alat, lääketieteen ja hoitotyön kirjallisuus, musiikkiaineistot, pedagogiikka sekä erilaiset kulttuurialojen aineistot. Lisäksi useiden erikoisalojen, hotelli- ja ravintola-alan, kauneudenhoito- ja turvallisuusalojen kirjallisuuden kuvailutietoja on tullut aineistonsiirroissa metatietovarantoon. Melindalle uudesta aineistosta pääosa on englanninkielistä. Uudentyyppisiä aineistoja kartunnassa ovat viittomakieliset aineistot ja elokuvat.