

LEHTITUETUEIDEN MATKASSA - digitoitujen lehtien metatietojen uudistaminen ja jakaminen Alma-yhteisölle

Turun yliopiston kirjastossa on käynnissä projekti Kotimaiset digitoituneet lehdet Community Zonelle. Tavoitteenamme on kohentaa Kansalliskirjaston digitoimien, kaikille avoimesti saatavilla olevien lehtien metatietoja. Metatiedot saatamme kirjastojen käytettäväksi paitsi kansallisesti, myös kansainvälisesti Alma-kirjastojärjestelmän metatietovarannon kautta. Vaikka olemme päivävuorossa projektista, matkan varrella on ollut mukana lukuisia muitakin tekijöitä.

Kaikki alkoi lehdestä nimeltä Borgåbladet. Miksi lehti oli tietokannassamme muodossa BORGAABLADET? Miksi sen bibliografiset tiedot olivat niin suppeat? Vuonna 2022 päätimme, että tiedot oli korjattava. Oletimme, että kyse ei olisi monimutkaisesta ja pitkällisestä urakasta. Pian kävi ilmi, että olimme väärässä.

Kotimaisten sanomalehtien tietueet muodostivat tietokannassamme sekavan kokonaisuuden. Painettujen lehtien osalta tilanne oli hyvä. Sen sijaan elektronisia versioita edustavien tietueiden kunto oli katastrofaalinen. Joukossa oli vuonna 2016 alas ajatun Nelli-portaalin yhteydessä muodos-

tettuja tietueita, joiden vähäinen kuvailutieto oli osittain virheellistä.

Kohtalokas hetki skandinaavisia merkkejä sisältävälle tietueille oli siirtyminen Almaan: nimeke katkesi näiden kohdalta. Osalle lehdistä kuvailutiedot oli poimittu Alma-kirjastojen yhteistietovarannosta, Community Zonesta (jatkossa CZ). Tietueiden taso ei päättä huimannut, esimerkkinä alussa mainittu Borgaabladet.

Lehtiä laajalle yleisölle ja yliopistoyhteisölle

Sanomalehtitietueita oli parannettava. Skandinaavisten merkkeihin puuttuminen ei nimittäin ollut ainoa asiakkaitamme hämmentävä asia. Tie-

Saatavuusvuodet	TY:n henkilökunta ja opiskelijat	Muut asiakkaat
1775–1939 (suomenkieliset) 1771–1949 (ruotsinkieliset)	Vapaasti verkossa	Vapaasti verkossa
–2021	Yliopiston tunnuksilla	Vapaakappaletyöasemilla
2022–	Vapaakappaletyöasemilla	Vapaakappaletyöasemilla

Digitoitujen lehtien saatavuus digi.kansalliskirjasto.fi-palvelussa.

Kuva: Kati Järvinen

tokantaamme sotkivat muun muassa tuplatietueet, vanhentuneet saatavuustiedot ja e-versiota edustavan tietueen puuttuminen.

Vapaakappalekirjastona tarjoamme painettuja ja digitoituja lehtiä niin yliopistoyhteisölle kuin laajemmalle yleisölle. Yliopiston tutkijat ja opiskelijat käyttävät sanomalehtiä tutkimuksessa, oppinäytetöissä ja erilaisilla kursseilla. Yliopistoyhteisön ulkopuoliset hyödyntävät lehtiä esimerkiksi lähdeaineistona tietokirjoissa, sukututkimuksessa sekä etsiessään tietoa menneistä tapahtumista.

Näiden moninaisten tarpeiden vuoksi olemme erityisen kiinnostuneita siitä, miten lehtien tiedot näkyvät tietokannassamme, ja miten tiedonhakija ne löytää. Painettuja sanomalehtiä on kokoelmisamme aina 1700-luvulta vuoden 2007 loppuun asti, jolloin vapaakappaleoikeutemme päättyi painettujen sanomalehtien osalta. Kotimaisia lehtiä on Kansalliskirjaston digitoinnin ansiosta saatavilla sekä yliopistoyhteisöllemme että muille asiakkaille.

Digitoidut lehdet ovat digi.kansalliskirjasto.fi-palvelussa. Tekijänoikeussyistä osa digitoiduista lehdistä on kuitenkin rajatussa käytössä. Asiakkaillemme onkin tarjolla digitoituja lehtiä kolmella eri tavalla, riippuen lehtien ilmestymisajankohdasta sekä asiakasryhmästä.

Ensimmäisen ryhmän muodostavat verkossa vapaasti saatavilla olevat digitoidut lehdet: suomenkieliset vuosilta 1775–1939 ja ruotsinkieliset vuosilta

LDR	----cas-a22----z-4500
008	230117uuuuuuuuuuuuuu- o---u --- swe-d
035	__ a (CKB)110978979778445
245	00 a BORGAABLADET
246	__ a Borgåbladet
260	__ b Tryckeri - och tidnings a Borgå
776	__ x 0358-6294
906	__ a NEWSPAPER

Borgåbladet-lehden tietue ennen.

1771–1949. Toinen ryhmä koostuu vuosina 1939–2021 ilmestyneistä lehdistä. Näiden lehtien ajasta ja paikasta riippumattomasta tutkimuskäytöstä olemme tehneet Tutkain-sopimuksen Kansalliskirjaston kanssa. Kolmas ryhmä ovat vuodesta 2022 eteenpäin ilmestyneet lehdet, jotka ovat käytössä ainoastaan vapaakappaletyöasemillamme.

Projekti käynnistyy kansallisella yhteistyöllä

Vuonna 2022 muissakin Alma-kirjastoissa koettiin tarpeelliseksi viedä suomalaisten aineistojen metatietoja CZ:lle. Syntyi kansallinen projekti Kotimaiset OA-aineistot CZ:lle, jossa Alma-kirjastojen kesken jaetaan käytäntöjä ja autetaan ratkomaan eteen tulevia ongelmia. Oman, digitoituihin lehtiin keskittyvän osaprojektimme nimesimme muotoon Kotimaiset digitoidut lehdet CZ:lle.

Ensimmäinen tehtävämme oli määrittellä projektiimme kuuluvat lehdet. Keskityimme vapaasti saatavilla olevien lehtien metatietoihin; olihan tavoit-

teenamme julkaista metatiedot myös CZ:lla. Ratkaistavaksemme tulee kuitenkin myös se, miten saamme omaan tietokantaamme näkyviin Tutkain-sopimuksen sekä vapaakappaletyöasemilla olevien lehtien saatavuustiedot.

Avoimesti saatavilla olevien sanomalehtien tietueita oli kahdentyyppisiä. Ensimmäinen joukko olivat ne noin 230 tietuetta, joilla oli painetun lehden bibliografinen tietue. Toisen joukon muodostivat ne noin 50 tietuetta, joilla oli elektronista vastaava tietue Melindassa.

Todistettavasti kiinnostusta tietueille on muuallakin kuin Suomessa, sillä eräs ruotsalainen yliopistokirjasto oli julkaissut suomalaisten lehtien tietueita CZ:lla.

Kuvailtavassa aineistossa oli eniten suomenkielisiä lehtiä (n. 60%), toiseksi eniten ruotsinkielisiä (n. 30%) sekä muutamia englannin- ja saksankielisiä lehtiä. Suomalaisen aineiston tarve Suomen ulkopuolella on toki rajallista. Todistettavasti kiinnostusta tietueille on kuitenkin muuallakin kuin Suomessa, sillä eräs ruotsalainen yliopistokirjasto oli julkaissut suomalaisten lehtien tietueita CZ:lla.

Pohdimme, miten saisimme digitoitujen lehtien tiedot näkyviin yksiselitteisesti asiakkaille. Jotta tilanne

korjaantuisi, tarvitsisimme kuvailusääntöjen mukaisesti omat tietueet elektronisen lehden manifestaatiolle. Kunnollisia e-versiota edustavia bibliografisia tietueita ei ollut.

Saatavuustiedot oli usein kytketty painetun tietueen yhteyteen, mikä ei ole kansainvälisten kuvailusääntösten mukaista. Alman CZ:lle ei haluta painettujen kuvailutietoja edustamaan e-versiota. Nykyisistäkin – joita on runsaasti – halutaan päästä eroon.

Ajatus luoda digitoiduille lehdille oma bibliografinen tietue sai varauksettoman tuen Kuvailevan metatiedon asiantuntijaryhmässä, Kumeassa. Kansalliskirjastossakin suhtauduttiin myönteisesti suunnitelmaamme, mikä olikin ehdoton edellytys projektin toteuttamiselle. Lehdethän ovat Kansalliskirjaston digitoimia, ja lisäksi olimme tekemässä muutoksia Fennica-tietueisiin.

Kuvailutietoja kohentamaan

Kansallisbibliografiapalveluissa tartuttiin toimeen. E-versioita edustavia tietueita ei ryhdytty laatimaan alusta pitäen vaan olemassa olevaa dataa hyödynnettiin niin paljon kuin mahdollista.

Melindasta poimittiin painettua aineistoa vastaavat bibliografiset tietueet niistä digitoiduista nimekkeistä, joilla ei ollut omaa e-tietuetta. Tietueista otettiin kopiot, ja ne muokattiin muunnosohjelmalla vastaamaan e-versiota. Tässä vaiheessa tietueet saivat ISSN-keskuksesta omat

ISSN:t, KKDIGIT-alkuisen merkin­nön 035-kenttään (=yhteinen nimit­täjä projektiin kuuluville tietueille) ja 588-kenttään tekstin ”Kuvailun perus­тана painetun aineiston tietue”.

Noin 50 tietueen muokkaus jäi mei­dän tehtäväksemme. Ne olivat verk­koversioina ilmestyneitä lehtiä, joille aikoinaan oli annettu oma e-ISSN. Kuvailun perusta piti muuttaa vastaa­maan varhaisinta varsinaista numeroa. Noudatimme vuodesta 2012 ollutta sääntöä (ISSN Network), jonka mu­kaan online-versiolle annettu ISSN edustaa myös digitoituja vuosikertoja. Tämä tarkoitti koko tietueen myllertä­mistä niin nimekkeen, julkaisu- ja nu­merointitietojen kuin muiltakin osin.

Mainitun kaltaista työtä, jossa digi­toituille lehdille annetaan takautuvasti ISSN, tehdään muuallakin, esimerkiksi Ranskan kansalliskirjastossa.

Tavoitteenamme oli saada tietueista laadukkaita ja voimassa olevan kuvai­lusäännön (=RDA) mukaisia. Eteen tulleet kuvailuongelmat toimitettiin Kumean sarja-aineistotyöryhmän rat­kaistavaksi. Ratkaisuisista työryhmä sai

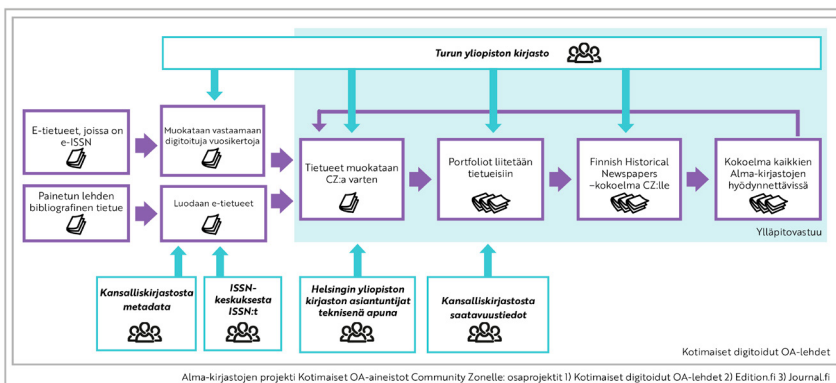
materiaalia työn alla olleeseen sarja-ai­neiston kuvailuohjeeseen. Ohjeessa on huomioitu seikkoja, jotka olisivat saat­thaneet jäädä vähäisemmälle huomiolle ilman projektiamme.

Tietueet kansainvälisesti saataville

S euraava vaihe oli saada tietueet kansainvälisesti saataville. Alma-kirjastona luonnollinen vaihtoehto tähän tarkoitukseen oli Ex-Libriksen (osa Clarivate-konsernia) Community Zone -tietovaranto. CZ on hyödyllinen, koska voimme sen avulla antaa ja ottaa käyttöön metada­taa koko Alma-yhteisön kesken. Yhteis­öön kuuluu n. 2 500 kirjastoa.

CZ:lla on bibliografista tietoa elektronisista aineistoista sekä niihin pääsystä kuten saatavuustiedoista ja resurssilinkeistä. ExLibris ei pidätä oi­keuksia itsellään siellä olevasta metatiedoista. Kirjastona meillä on täysi oikeus CZ:lle tuottamaamme metadataan: se on käytössämme senkin jälkeen, kun emme enää ole Alma-järjestelmässä.

Käytännössä voimme ladata datan



Kotimaiset digitoidut lehdet CZ:lle -projektin kulku. Kuva: Päivi Letonsaari

itsellemme. Metadata jää kuitenkin CZ:aan ExLibriksen kirjastoasiakkaiden käyttöön, vaikka vaihtaisimme järjestelmätoimittajaa. Luovumme siis kaikista oikeuksista.

CZ:lla on tietyt säännöt, joita tietueiden tulee noudattaa. Tietueista poistetaan Melinda-kytkös. Mahdollisia Melinda-päivityksiä varten meillä on tietokannassamme erilliset tietueet, joissa on yhteys Melindaan.

Uusia työkaluja tietueiden muokkaamiseen

Seuavaa vaihe oli muokata 230 tietuetta CZ:lle kelpaaviksi. Kyseessä olivat niiden sanomalehtien bibliografiset tietueet, joiden kuvailu perustui painetun aineiston tietueisiin. Tämän vaiheen myötä pääsimme tutustumaan kahteen uuteen työkaluun.

Ensimmäinen meille uusi työkalu oli MarcEdit-ohjelmisto. Vinkin siitä sekä käyttöopastusta saimme Helsingin yliopiston kirjaston asiantuntijoilta. MarcEdit nopeutti työtämme, sillä pystyimme muokkaamaan suuren määrän tietueita samalla kertaa. Esimerkiksi 856-linkkikentät sai poistettua kerralla kaikista tietueista.

Toinen meille uusi asia oli CZ:n tietueille asettamat vaatimukset. CZ:lla kuvailukielen on oltava joko aineiston kieli tai englanti. Siksi muutimme ruotsinkielisten lehtien kuvailutiedot ruotsiksi sekä saksan- ja englanninkielisten lehtien kuvailutiedot englanniksi. Käännettäviä kenttiä olivat esimer-

kiksi 338-kentän ”verkkoaineisto” ja 588-kentän ”Kuvailun perustana painetun aineiston tietue”.

Joitain tietueita oli vielä muokattava yksitellen ennen kuin ne olivat soveltuvia CZ:lle. Erilaisia tekstivariaatioita saattoi olla useita, joten niitä ei voinut korvata kaikkia samalla fraasilla. Esimerkiksi lehden ilmestymistiheydestä kertovia käännöksiä tarvittiin lukuisia.

Portfoliot paljastavat aineistojen salaisuudet

Seuuraavaksi tarvitsimme portfoliot, joista asiakkaille selviävät lehtien saatavuustiedot. Itse resurssit eli julkaisujen digitoitu data on Kansalliskirjaston palvelussa. Kansalliskirjasto tarjoaa helposti ladattavat saatavuustiedot Excel-tiedostona. Excelin dataa hyödyntämällä loimme kirjastojärjestelmään jokaiselle tietueelle portfolion, johon liitimme julkaisun saatavuusvuodet sekä asianmukaisen linkin Kansalliskirjaston aineistoihin.

Saatavuustietoihin jäi kuitenkin avoimia kysymyksiä. Ensinnäkin digitoitujen lehtien saatavuudet ovat erilaisia eri asiakasryhmille. Erojen tekeminen selväksi asiakkaille ilman väärinymmärryksiä on suuri haaste. Parhaimman tavan miettiminen onkin meillä vielä edessä.

Toinen ongelma on, että saatavuustiedot päivittyvät jatkuvasti digitoinnin edistyessä. Lisäksi aineiston vapaa saatavuus muuttunee vuosien kuluessa, koska tekijänoikeudellisista rajoituksista neuvotellaan säännöllisesti.



Toistaiseksi olemme suunnitelleet päivittävämme saatavuustiedot vuosittain.

Merkittävä virstanpylväs sekä jälleen yllätyksiä

Projektimme eteni vuoden 2023 marraskuussa merkittävään vaiheeseen, kun saimme julkaistua Finnish Historical Newspapers -kokoelman CZ:lla. Tilanne oli käänteentekevä. Olimme aiemmin aktivoineet aineistoja CZ:lta, mutta nyt osallistuimme itse myös tietovarannon rikastamiseen.

Yhteistietovarantojen etuna on kirjaimellisesti yhteistyö – kaikkea ei tarvitse tehdä tai tuottaa itse. Vastavuoroisesti niiden haittapuolena on tietueiden laadullinen kirjo ja erilaiset laatutasot kuvailusäännöistä sekä -normeista huolimatta, ja tämä näkyy

myös CZ:lla. Tietovarannon rosoisuus – tietueiden kirjavuus ja tuplatietueet – tuli eteemme, kun vastuullemme jäi kokoelman ylläpito, seuranta ja päivittäminen.

Ylläpitoon kuuluu meille yllätyksenä tullut tehtävä. Meidän pitää valvoa kokoelmamme bibliografisten tietueiden tilannetta CZ:lla aina kun päivitämme kokoelmaa. Kokoelman ensimmäisen julkaisemisen yhteydessä osa portfolioista yhdistyi väärin bibliografisiin tietueisiin. Näin käy, jos nimikkeistä on jo valmiiksi tietue, jolla on sama ISSN. Harmiksemme nämä muutamat väärin yhdistyneet tietueet olivat tasoltaan varsin heikkoja – kuten esittelemämme Borgiaabladet tekstin alussa.

Väärin yhdistymisten korjaaminen vaati yhteydenottoa ExLibrikseen

Kotimaiset digitoidut lehdet CZ:lle -projekti

Kotimaiset digitoidut lehdet Community Zonelle -projektin päävastuu on Turun yliopiston kirjastolla. Projekti on yksi Alma-kirjastojen projekteista "Kotimaiset OA-aineistot Community Zonelle". Muita projekteja ovat journal.fi ja edition.fi.

Artikkelin kirjoittajien lisäksi projekti on koskettanut - enemmän tai vähemmän - seuraavia tahoja ja henkilöitä

- Kansalliskirjaston kausijulkaisujen kuvailijat
- Heidi Ronkainen, ISSN-keskus
- Maria Kovero ja Ville Huhtala, Helsingin yliopiston kirjasto
- Sarja-aineistotyöryhmä (Kumean alatyöryhmä)
- ExLibriksen Community Zone -asiantuntijat.

LDR	01896cas a2200445 4500
001	99110978979778445
005	20240423120421.0
007	cr#
008	220519d18602015f dr n o b0swe
022	_ a 2814-5496
035	_ a (CKB)110978979778445
040	_ a FI-NL b swe e rda d FI-Hu d FI-T
041	0_ a swe
043	_ a e-fi--
044	_ a fi
050	_4 a AN
222	0_ a Borgåbladet b (Online)
245	00 a Borgåbladet : b tidning för Borgåstad och socken, Sibbo och östra Nyland
264	_1 a Borgå : b Tryckeri- och tidnings, c 1860-2015.
300	_ a 1 onlineresurs
310	_ a Sex gånger per vecka
321	_ a Fem gånger per vecka
336	_ a text b txt 2 rdacontent
337	_ a dator b c 2 rdamedia
338	_ a onlineresurs b cr 2 rdacarrier
362	0_ a N o 0 (lördagen den 15 december 1860)-vecka 2, årgång 153, nr 5 (lördag 10 januari 2015).
506	0_ 3 1860-1949 a Materialet är fritt tillgängligt. f Unrestricted online access 2 star
550	_ a Senare utgivare: KSF Media.
588	_ a Beskrivningen baserad på det tryckta materialets katalogpost.
651	_7 a Suomi. 0 http://www.yso.fi/onto/yso/p94426 2 yso/fin
651	_7 a Finland. 0 http://www.yso.fi/onto/yso/p94426 2 yso/swe
655	_7 a paikallislehdet. 0 http://urn.fi/URN:NBN:fi:au:sim:s224 2 sim/fin
655	_7 a kausijulkaisut. 0 http://urn.fi/URN:NBN:fi:au:sim:s1332 2 sim/fin
655	_7 a sanomalehdet. 0 http://urn.fi/URN:NBN:fi:au:sim:s39 0 http://urn.fi/URN:NBN:fi:au:sim:s39 2 sim/fin
655	_7 a lokaltidningar. 0 http://urn.fi/URN:NBN:fi:au:sim:s224 0 http://urn.fi/URN:NBN:fi:au:sim:s224 2 sim/swe
655	_7 a periodika. 0 http://urn.fi/URN:NBN:fi:au:sim:s1332 2 sim/swe

Borgåbladet-lehden tietue nyt.

tai siihen kirjastoon, joka oli tuottanut alkuperäisen tietueen. CZ:n tietueissa ISSN-kenttä on lukittu. Jotta muutoksia näihin voitiin tehdä, piti ExLibriksen ensin muuttaa CZ:lla oleva ISSN-kenttä vastaamaan Kansalliskirjaston antamaa uutta e-ISSN:ää. Onneksi edellisen kaltaisia korjausta vaativia tapauksia oli häviävän pieni osa tietueistamme. Niihin kuitenkin kuluu aikaa ja resursseja.

Ylläpitoon kuuluu myös kokoelmasta vielä puuttuvien lehtien lisääminen. Puuttuvat lehdet voivat olla Kansalliskirjaston kokonaan uusina digitoimia. Lisäksi kokoelmasta puuttuu

kirjoittamishetkellä sellaisia vielä ilmesytyviä lehtiä, joiden kuvailun lähtökohta on ollut moderni uutisvirtalehti. Nämä tietueet on ulotettu kattamaan myös lehden vanhat digitoituid vuosikerrat.

Työ jatkuu

Vielä noin 50 sanomalehtitietuetta tarvitsee viimeistelyä ennen kuin ne ovat mukana Finnish Historical Newspapers-kokoelmassa. Jatkamme projektiamme Kotimaiset digitoituid lehdet CZ:lle aikakauslehtien parissa. Tulevaisuudessa aikakauslehtitietueet muodostavat oman kokoelman, Finnish Historical Journals.

Tällä kertaa osaamme jo aavistaa, mihin olemme ryhtymässä. Lisäksi olemme oppineet prosessin aikana monenlaista: MarcEditin käyttöä, kausijulkaisujen kuvailuun liittyviä yksityiskohtia, kirjastojärjestelmämme ominaisuuksia ja yhteistyötä ExLibriksen kanssa. Ennen CZ oli tullut meille tutuksi poiminnan lähteenä, nyt saimme kokemusta mitä kaikkea liittyy siihen, kun CZ:lla ollaan liikkeellä oman kokoelman kanssa.

Alussa mainittu katastrofitietuekin on saatu korjattua. Vuonna 2024 asiakkaamme voivat löytää sen nimellä Borgåbladet: tidning för Borgåstad och socken, Sibbo och östra Nyland. ❖

Kirjoittajat

ANNE-MARI SALMELA
Turun yliopiston kirjasto
anne-mari.salmela@utu.fi

JOUNI AALTONEN
Turun yliopiston kirjasto
jouni.aaltonen@utu.fi

SARA KANNISTO
Turun yliopiston kirjasto
sara.kannisto@utu.fi

ANTTI MALIN
Turun yliopiston kirjasto
anvama@utu.fi