

Dublin Coren oppivuodet

Juha Hakala

Informaatioteknologian näkökulmasta metadataformaattien historia on pitkä. Kirjastoammattilaiset ovat tuottaneet konelukuisia kuvailutietoja jo 70-luvulta lähtien, ja vieläpä samassa formaatissa. Vasta viimeisten 10 vuoden aikana metadatasta on tullut muotia: sitä tuotetaan ja käytetään moniin eri tarkoituksiin, monilla eri aloilla ja useissa eri formaateissa. Tämä artikkeli kuvaa Dublin Core -metadataformaatin kehitystä. Sain seurata läheltä tämän metadataformaatin kehitystä 15 vuoden ajan ja omalta osaltani olla myös sen kehitystyössä mukana.

Metadataformaattien historian tutkiminen voi olla vaikeaa. Keskeisetkin päätökset jäävät joskus dokumentoimatta. On vaikeaa selvittää esimerkiksi sitä, miksi MARC-formaatin koeversiossa oli kiinteämittaiset kentät, ja mitkä tekijät vaikuttivat siihen, että ne ensimmäisessä tuotantoversiossa korvattiin vaihtuvamittaisilla kentillä, joiden pituus ilmoitettiin tietueen nimiössä. MARCin kehittämisen aikaan tämä oli hyvin innovatiivinen ja poikkeuksellinen ratkaisu. Ja kallis, koska vaihtuvamittaiset tietueet saattoivat 70-luvun katsannossa viedä paljon tilaa.

2000-luvulla metadataformaatin kehittäjien ongelmat eivät enää liity laitteistojen rajoitukseen vaan formaattien soveltamiseen. Dublin Coren tarina on tästä hyvä esimerkki. Se poikkeaa yhdessä suhteessa oleellisesti kaikista edeltäjistään, kuten kirjastojen MARC-formaatista. Dublin Core ei ole sidottu mihinkään tiettyyn ympäristöön, vaan sitä voidaan soveltaa periaatteessa missä vain. Tämä abstrakti ja vaikeasti saatutettava lähtökohta muotoiltiin kunnianhimoiseksi tavoitteeksi määritellä kuvailevan metadatan semanttinen ydin, ”core”.

Formaatin synty

OCLC ja National Center for Supercomputing Applications (NCSA) järjestivät 1.-3 maaliskuuta 1995 OCLC:n tiloissa Dublinissa, Ohiossa se-

minaarin, jossa pohdittiin sitä, voitaisiinko yksinkertaisen metadataformaatin avulla kuvailla laajaa kirjoa erilaisia Internetin tiedonlähteitä. Taus-ta tälle seminaarille oli Stu Weibelin mukaan ensimmäisessä WWW-konferenssissa kesäkuussa 1994 kahvittauolla käyty keskustelu, jossa pohdittiin sitä, miten jo tuolloin räjähdysmäisesti kasvavan Internetin tiedonhaun ongelmia voitaisiin tehokkaasti ratkoa. Korkealaatuisen metadatan avulla, oli keskustelijoiden yksimielinen vastaus.

Vuoden 1995 ensimmäinen Dublin Core workshop ja sitä seuranneet neljä DC-kokousta on dokumentoitu D-Lib magazine -lehdessä julkaistuissa artikkeleissa (Weibel95, Weibel96, Weibel97a, Weibel97b ja Weibel98). Yhdessä nämä tekstit kuvaavat sekä suoraan että rivien välissä sitä, miten metadataformaatti ja sitä kehittävä yhteisö rakentuvat yrityksen ja erehdyksen kautta.

Kuvaus ei välttämättä ole yleispätevä, koska formaatit kehitetään yleensä jollekin sektorille sen erityistarpeet huomioon ottaen. Hyviä esimerkkejä tästä ovat kirjastojen MARC-formaatti ja kirjakauppa-alan kehittämä ONIX-formaatti, jotka kuvaavat samoja bibliografisia objekteja, mutta hyvin eri näkökulmista. ONIXissa on tärkeää kuvata esim. kirjan mitat, jotta kirjakauppa voi arvioida, miten monta nidettä uutuushyllyyn mahtuu. Dublin Core puolestaan pyrkii nimensä mukaisesti olemaan semanttinen ydin, eräänlai-

nen eri metadataformaattien yhteinen nimittäjä. Tästä 15 kentän kokoisesta ytimestä on vuosien mittaan kehitetty sovellusprofiileja, eli versioita erilaisiin käytännön tarpeisiin, kuten opinnäytteiden tai asiakirjojen kuvailuun.

Tiedonhaun tuki tavoitteena

Vuonna 1995 ei tulevasta kehityksestä ollut vielä kuitenkaan aavistusta. Ensimmäisen DC-workshopin osallistujien keskeisenä huolena oli rajata tarvittavien metadataelementtien (kenttien) määrä. Päätettiin, että Dublin Corella piti ensisijaisesti pystyä kuvaamaan dokumentin kaltaisia objekteja (document-like objects) ja että kuvailun elementtien piti tukea tiedonhakua (discovery).

Jälkiviisaasti voi sanoa, että molemmat rajaukset olivat enemmän haitaksi kuin hyödyksi. DC-yhteisö ei koskaan saavuttanut yhteisymmärrystä siitä, mitä dokumentin kaltaisella objektilla oikeastaan tarkoitetaan. Onneksi tämä rajaus ei juuri sitonut soveltajien käsiä. Dublin Coresta tehtiin tuota pikaa sovelluksia esimerkiksi kokoelmien ja muiden varsin niukasti dokumenttien kaltaisten objektien kuvailuun.

Tiedonhakua koskeva rajaus sitä vastoin oli käytännössä ongelmallinen. Linjauksesta tuli allekirjoittaneen mielestä haitallinen viimeistään siinä vaiheessa, kun Dublin Coren käyttäjien kannalta järkeviä ehdotuksia uusiksi termeiksi hylättiin, sillä perusteella, etteivät ne Usage boardin mielestä tukeneet tiedonhakua. Ehdotuksen tekijät saattoivat täysin perustellusti olla asiasta päinvastaista mieltä. Usage board -käsittelyn lopputuloksen ennakoimisesta tuli myös vaikeaa, koska on pitkälti makuasia, onko jokin tietoelementti haun kannalta hyödyllinen vai ei.

Globaali suuntautuminen

Ironista kyllä, vuoden 1995 DC-seminaaria kuvaavan artikkelin nimi on Metadata: The Foundations of Resource Description. Kun Dublin Coren tavoitteita mietittiin johtoryhmän retriittisä vuonna 2003, hyväksyttiin se ajatus, että formaatti pyrkii edistämään resurssien (eikä siis do-

kumentin kaltaisten objektien) kuvailua, riippumatta siitä oliko formaatin metadataelementeillä tekemistä haun kanssa eli ei. Tämän jälkeen Dublin Coren soveltamisala on entisestään laajentunut.

Allekirjoittanutta huvittava päätepieste formaatin tavoitteen asettamiselle on se keskustelu, jota käytiin syksyllä 2011 vuoden 2012 Dublin Core-konferenssin nimestä: vahva kandidaatti oli Metadata for solving global problems. DC-väkeä voisi syyttää megalomaanisuudesta, mutta minusta tässä on kyse pikemminkin siitä, miten metadatan rooli Internetissä yleensä on kasvanut viimeisten 15 vuoden aikana.

Vaikka tehdyistä rajauksista voidaan olla montaa mieltä, ensimmäisen seminaarin osallistujat loivat Dublin Coren semanttisen ytimen. Kenttiä ei tosin ollut viittätoista vaan kolmetoista - Rights ja Description lisättiin vasta myöhemmin, kolmannen DC-tapaamisen jälkeen. Mutta monia keskeisiä asioita lykättiin alkuvaiheessa tuonemmaksi, kuten DC-metadatan esittäminen koneluettavassa muodossa (syntaksi), formaatin käyttökohteiden määrittely (pragmatiikka) sekä Dublin Coren ylläpidon ja kehittämisen organisointi. Näiden asioiden parissa jouduttiin tekemään työtä vielä useita vuosia.

Dublin Coren tietoelementit

Dublin Coren standardoiduissa versioissa (ISO, SFS) on 15 tietoelementtiä (kenttää). Paljon on keskusteltu siitä, ovatko ne todellakin varsinainen ydin. Kritiikin aiheita on ollut monia. Esimerkiksi kirja-alan toiveena oli, että Creator ja Contributor sekä mahdollisesti Publisher yhdistettäisiin kenttään Agent, jossa rooli-indikaattorin avulla osoitettaisiin tekijyyden luonne. Lisäksi Source ja Relation on haluttu yhdistää, koska lähteenä toimiminen on eräänlainen suhde sekkin. Mutta formaatin ydin, perinteinen Dublin Core, on sitä, mitä se on, eikä sitä tulla enää muuttamaan.

Formaattia huonosti tunteva voi luulla, että 15 kenttää ovat se, mitä Dublin Core tarjoaa. Mutta kuten jokainen ahkerasti käytetty metadatafor-

maatti, Dublin Core on kehittynyt jatkuvasti. Jo 90-luvun puolella kentille laadittiin tarkenteita (Qualifiers). Erityistä vaivaa nähtiin Date- ja Relation-kenttien kanssa. Esimerkiksi edellisen tarkenteet olivat Created, Valid, Available, Issued ja Modified (Weibel 00). Myöhemmin kehitettiin sovellusprofiileja erityistarkoituksiin, kuten kokoelmien kuvailuun. Niiden sisältämät termit eivät enää ole ydinkenttien tarkenteita, vaan aitoja semanttisia laajennuksia formaattiin.

Toki Dublin Core on edelleen muihin laajasti käytettyihin formaatteihin verrattuna suppea: MARCissa on reilut 200 kenttää ja ONIXissa peräti 500, ja tietoelementtejä vielä paljon enemmän. Dublin Coren termejä (<http://dublincore.org/documents/dcmi-terms/>) on tätä kirjoitettaessa vasta vähän yli 50. Sovelluskohteiden vuoksi muutamat termit ovat kuitenkin varsin spesifejä. Esimerkiksi kokoelmasta voidaan kuvata kokomapolitiikka ja aineiston hankintamenetelmä.

Toimialariippumattomuus ja erityissovellukset

Nyky muodossaan Dublin Core ei ole enää vain semanttinen ydin, joka sellaisenaan ei sovellu tarkkaan resurssien kuvailuun, vaan formaattia on laajennettu tiettyihin erityistarpeisiin soveltuvaksi. Pohjimmiltaan formaatin tavoitteet ovat kuitenkin samat kuin ne olivat 15 vuotta sitten (Weibel96):

1. A simple interchange format for descriptive metadata
2. Content self-description for networked objects
3. Semantic interoperability across domains

Mikään näistä tavoitteista ei ole vanhentunut, vaan pikemminkin päinvastoin. Semanttinen Web ei toimi ilman semanttista yhteismitallisuutta, joka on myös Suomessa yksi julkishallinnon tietoarkkitehtuurin keskeisistä tavoitteista. Dublin Core onkin ainoa julkishallinnon standardisalkussa (<https://www.yhteentoimivuus.fi/>) mainittu metadataformaatti, koska toimialakohtaisia formaatteja kuten kirjastojen MARCia ei otettu salkkuun mukaan.

15 vuotta sitten ei kuitenkaan vielä nähty sitä, miten suuri tarve eri projekteilla oli soveltaa formaattia omiin tarpeisiinsa, ja siinä sivussa laajentaa Dublin Corea. Jos minulle olisi 1996 kerrottu että 15 vuotta myöhemmin formaatista laaditaan Kansalliskirjastolle versio keskiaikaisten käsikirjoitusfragmenttien kuvailuun, en luultavasti olisi uskonut moista.

Jos Dublin Core -kehityksessä mukana olevilla oli vaikeuksia hahmottaa mistä oli kyse, vielä suurempia vaikeuksia oli niillä jotka seurasivat asiaa sivusta. ”Dublin Core on vain huonosti tehty MARC -formaatti” oli takavuosina tavallinen kommentti. Jos on sisäistänyt vain omalle toimialalle optimoidun formaatin ja sen soveltamisessa käytettävät luettelointisäännöt, voi olla vaikeaa ymmärtää paitsi muiden toimialojen kuvailukäytänteitä, myös Dublin Coren kaltaista toimialariippumatonta ratkaisua. Nykyään Dublin Coren tarpeellisuutta harvemmin asetetaan kyseenalaiseksi. Semanttinen Web, Dublin Coren tehtäväkohtaiset sovellukset kuten DC Collections Application Profile (<http://www.dublincore.org/groups/collections/collection-application-profile/>) sekä Dublin Core -tuki Microsoft Officen kaltaisissa tuotteissa ovat osoituksia formaatin nykystatuksesta ja tarpeellisuudesta.

Syntaksi

Metadataformaateissa pysyvintä on semantiikka, koska se usein perustuu kuvailusääntöihin, joiden radikaaliin muuttamiseen ei millään yhteisöllä ole varaa. Vuosikymmenien mittaan luodut kuvailutiedot ovat perusta, joka on otettava huomioon, vaikka kirjastoalalla kuvailusääntöiksi tuleekin RDA, ja MARCin korvaa ennen pitkää jokin muu formaatti.

MARC on mielenkiintoinen tapaus sen vuoksi, että sen syntaksi ei ole muuttunut miksiäkään 40 vuoteen, sen jälkeen kun formaatista otettiin käyttöön vaihtuvamittaiset kentät salliva versio. Alkuun MARC oli teknisesti edistynyt formaatti, mutta vuosikymmenien mittaan siitä on tullut reliikki, jota jokainen itseään kunnioittava alan

nuori ammattilainen mätkii menen tullen. Mitään ei kuitenkaan pidä rikkoa ennen kuin korvaava ratkaisu on valmis.

MARCiin verrattuna Dublin Coren syntaksiratkaisun kehittäminen on ollut melkoista vuoristorataa, koska Internet ja Web ovat tältä osin edelleen ”work in progress”. Kyse on siitä, missä muodossa avointa linkitettyä dataa ryhdytään julkaisemaan.

Tallennusalustat ja haettavuus

Ensimmäiset käytännön Dublin Core -hankkeet kuten Nordic Metadata Project rakensivat tallennusalustoja, joiden avulla käyttäjät pystyivät tallentamaan kuvailutiedot HTML:n meta-kenttään. Vastaavanlaisia tallennusalustoja on käytössä yhä, stand-alone -sovelluksina esimerkiksi opinnäytteiden kuvailujen tallennukseen, tai sisäänrakennettuna esimerkiksi Office-sovelluksiin. Dublin Core -yhteisö on luonut ohjeita sille, miten metatietoja tulisi tallentaa dokumentteihin (HTML:n ja XHTML:n Meta-kentän osalta katso <http://dublincore.org/documents/dc-html/>).

Dublin Core -metadatan esittämisen ohella Dublin Core -yhteisö pyrki kehittämään teknisen ratkaisun erilaisten metatietojen tallennukselle samaan ”säiliöön” (Lagoze). Ainakaan alikirjoittanut ei täysin sisäistänyt mistä puhuttiin, mutta ensimmäiset käytännön sovellukset teemasta ilmestyivätkin vasta muutamia vuosia myöhemmin. METS (<http://www.loc.gov/standards/mets/>), WARC ja muut säiliöstandardit ovat nykyään arkipäivää esimerkiksi digitaalisen arkiston hankkeissa.

90-luvulla ongelmana oli kuvailutietojen laatu - hakurobotit lopettivat tuolloin META-kentän indeksoinnin, koska metadatalle usein pyrittiin robottien yksinkertaisten ranking-algoritmien huijaamiseen. Tallentamalla META-kenttään sama sana tuhansia kertoja tai kymmeniä tuhansia eri sanoja pyrittiin sivuille saamaan korkea ranking.

Kaiken kaikkiaan korkeatasoista metadataa ei ollut riittävästi. Dublin Core -yhteisö toki yrit-

ti parhaansa: ohje Dublin Core -metadatan HTML-koodauksesta julkistettiin jo 1999 (<http://tools.ietf.org/html/rfc2731>), ja sitä seurasi kaksi RDF-ohjetta 2002. Viimeisin syntaksimäärittäminen (tätä kirjoitettaessa vuodelta 2008) löytyy osoitteesta <http://dublincore.org/documents/dc-rdf/>. DC-muotoisen metadatan tuottamisen konelukuisessa muodossa pitäisi siis olla helppoa.

Google ja Schema.org

Liekö sitten Dublin Coren ansiota että myös hakukoneiden kehittäjät uskovat metadatan voivan parantaa kokotekstihaun tulosta. Ne ovat Googlen johdolla kehittäneet oman metadataratkaisun nimeltä Schema.org (<http://schema.org/>) joka tarjoaa semantiikan ja syntaksin jota on sovellettava, jos haluaa saada metatietonsa Googlessa haettaviksi. Tämä politiikan muutos kertoo siitä, että hakukoneiden salaisissa ranking-järjestelmissä metadatan käsittely on niin edistynyt, etteivät alkeelliset huijausyritykset enää onnistu. Kun sosiaalisen metadatan tuottamiseen on totuttu, hakukoneet voivat saada niin suuria määriä luotettavaa metadataa, että ne parantavat kokoteksti-indeksoinnin tulosta. Tämän parempaa todistusta metadatan merkityksestä me tuskin voimme saada.

Mutta Schema.org on haaste perinteisille toimijoille, kahdella tapaa. Se tarjoaa resurssien kuvailuun oman semantiikkarakkaisunsa (ks. esimerkiksi monografioiden kuvailun elementit osoitteesta <http://schema.org/Book>) joka ei edes pyri olemaan yhteismitallinen Dublin Coren tai muiden perinteisten metadataformaattien kanssa. DC-yhteisössä ollaan oltu kahta mieltä siitä, pitääkö luoda mappaus DC-elementeistä Schema.org -elementteihin. Jotkut näkevät, että tässä annetaan pikkusormi pirulle. Ennen pitkää voidaan olla tilanteessa jossa jotkut jo kysyvät, tarvitaanko Dublin Corea enää.

Ennen kuin vastaa tähän kannattaa miettiä haluammeko me antaa tämäntyyppisiä vastuita Googlen kaltaiselle toimijalle tai laajemmin yksityiselle sektorille. Tilanne on vähän sama kuin

jos korvaisimme MARCin ONIXilla, joka on kehitetty erityisesti kirja-alan tarpeisiin mutta ei sellaisenaan käy kirjastolle. Schema.org on rakennettu ajatellen Internetin hakupalveluita yleensä ja Googlea erityisesti, mutta Dublin Coren soveltamisala on tätä laajempi.

RDF ja konversioiden tarve

Semantiikan ohella toinen Schema.org'in herättämä periaatteellinen kysymys koskee syntaksia. W3C on jo yli kymmenen vuoden ajan markkinoinut ja kehittänyt edelleen Ora Lassilan luomaa Resource Description Framework -ratkaisua (<http://www.w3.org/RDF/>). Siihen perustuu RDFa (<http://www.w3.org/TR/xhtml-rdfa-primer/>), Dublin Coreen nojaava menetelmä metatietojen tallentamiseksi XML-julkaisuun. RDFa mahdollistaa ainakin periaatteessa sen, että esim. verkkojulkaisun tiedot saadaan koneymärrettäviksi.

Schema.org ei sovelta RDF:ää, vaan vaatii oman XML-pohjaisen syntaksin. Yleisellä tasolla tämän voi nähdä haasteena W3C:lle ja sen versiolle semanttisesta webistä. Dublin Coren käyttäjille Googlen toiminta merkitsee sitä, että metatiedot on muunnettava eri syntaksiin riippuen siitä, minne ne toimitetaan.

Näissä konversioissa ei sinänsä ole mitään uutta, niitähän joudutaan tekemään aina kun MARC-dataa toimitetaan esimerkiksi digitaalisen aineiston hallintasovelluksiin tai asiakasliittymiin. Mutta Schema.org on periaatetasolla ikävä esimerkki siitä, miten suuret toimijat sanelevat omat pelisääntönsä joihin muiden on sopeuduttava. Tämä edesauttaa Webin ja laajemmin Internetin fragmentoitumista. On syntynyt Facebookin kaltaisia erillisiä saarekkeita, joissa olevaa (meta)tietoa voi käyttää vain kirjautumalla kyseiseen palveluun. Tämä on epäilemättä palveluiden omistajien etu, ja aivan yhtä varmasti Internetin käyttäjien ongelma.

Formaatin ylläpito-organisaatio

Kun metadataformaattia on kehitetty 15 vuot-

ta, sen taustalla on pakko olla jokin enemmän tai vähemmän pysyvä organisaatio. Dublin Coren osalta tämä organisoitumisvaihe vei muutamia vuosia, mutta vuonna 2011 toiminta on jo hyvin vakiintunutta.

Dublin Core Metadata Initiativen alkuvuosina formaatti oli riippuvainen Stuart Weibelin henkilökohtaisesta panoksesta sekä hänen työnantajansa OCLC:n taloudellisesta tuesta. Jotkut vetivät tästä sen johtopäätöksen, että OCLC vaikutti myös formaatin sisältöön, mutta itse en havainnut tästä koskaan mitään merkkejä. Vaikka OCLC:n johto tuki formaattia taloudellisesti, se jääväsi itsensä aina kaikista sisältökysymyksistä.

Ensimmäisissä vuosissa formaatin kehittämisessä oli huutoäänestyksen piirteitä. Esimerkiksi Helsingin kokouksessa 1997 (Weibel98) ja sitä seuranneessa Washington DC:n kokouksessa kukaan ei voinut olla täysin varma siitä, lähteekö projekti lainkaan siivilleen. Pohjanoteeraus koettiin mielestäni Washingtonissa, missä kehittämissyhteisö oli revetä kahtia. Kirja-alan edustajat halusivat formaattiin merkittäviä muutoksia, kuten Creator- ja Contributor-kenttien yhdistämistä. Tämä Agent proposalin nimellä kulkenut ehdotus on edelleen asia, josta on verisimmin yhteisön sisällä riideltä. Kun enemmistö oli sitä mieltä, että kentät jäävät sellaisekseen, kirja-alan väki ryhtyi kehittämään ONIX-formaattia. Jälkiviisaasti tapahtunutta voi pitää hyvänä asiana, koska yksi hyvin suuri käyttäjäyhteisö olisi voinut ”kaapata” formaatin itselleen.

Jo 10 vuotta sitten formaatin kehittäminen oli vakiintunut pitkälle nykyisiin uomiinsa. Tuon aikainen organisaatio (<http://dublincore.org/documents/dcmi-structure/>) on osin edelleen olemassa, vaikka vastuuhenkilöt ovat vaihtuneet.

Työryhmät kehitystyön takana

Nykyisessä DCMI:ssä on enemmän tai vähemmän pysyviä yhteisöjä (communities) kuten DC Libraries tai DC Government, sekä periaatteessa määräaikaista työryhmiä kuten DCMI/RDA tai DCMI Registry. Nämä ryhmät koostuvat va-

paahtoista, ja ryhmien aktiivisuus on viime kädessä paljolti puheenjohtajasta kiinni. Vuosien mittaan työryhmiä on tullut ja mennyt, ja ovat muutamat yhteisötkin vaipuneet unholaan.

Työryhmien ja yhteisöjen lisäksi keskeisiä toimielimiä ovat 2000-luvun alussa perustettu Usage Board, joka mm. hyväksyy uudet termit ja käsittelee ehdotukset uusiksi sovellusprofiileiksi, sekä Oversight committee eli johtokunta (<http://dublincore.org/about/trustees/>), jonka jäsenet edustavat niitä maita / organisaatioita, jotka ovat liittyneet DCMI:in. Suomi oli ensimmäinen jäsenvaltio, ja meitä ovat seuranneet muun muassa Uusi Seelanti, Iso-Britannia, Etelä-Korea ja viimeksi Malesia. Suomen edustaja johtokunnassa on syyskuusta 2011 eteenpäin Ulla Ikäheimo Kansalliskirjastosta. Ullan edeltäjä oli allekirjoittanut.

DCMI:n tulot koostuvat jäsenmaksuista ja satunnaisista tuloista kuten konferenssien osallistumismaksuista, menot ovat etupäässä palkkoja. DCMI Executive (<http://dublincore.org/about/executive/>) koostuu DCMI:n palkkalistoilla olevista henkilöistä. Heillä, ja ennen kaikkea Chief Executive Officerilla (Stuart Sutton), on päävastuu siitä, että formaatti kehittyä ja voi hyvin. Stuartilla, kuten kaikilla muillakin Executiven jäsenillä on taustallaan pitkä DC-ura vapaaehtoisena. Stu Weibel, johon Dublin Core alkuvaiheessa pitkälti henkilöityi, jätti Executiven jo lähes 10 vuotta sitten. Hänen jälkeensä ruorissa oli Max Dekkers, joka puolestaan väistyi johdosta vuoden 2011 heinäkuussa.

Kehitystyö jatkuu

Dublin Core on ainoa metadatastandardi, jota voidaan soveltaa periaatteessa kaikenlaiseen aineistoon. Tämän joustavuuden kääntöpuoli on se, että kukaan ei tiedä, ketkä ja mihin tarkoituksiin ja millaisilla välineillä formaattia käytetään tulevaisuudessa.

Kansalliskirjasto, joka vastaa Dublin Coresta kansallisesti, on perillä ensisijaisesti vain kirjastosektorin hankkeista. Julkishallinnon hankkeis-

ta merkittävä osa on Arkistolaitoksen vastuulla, ja yksityisen sektorin projektit hoituvat yleensä yritysten omilla resursseilla. Kukaan ei tiedä kattavasti miten Dublin Core Suomessa sovelletaan. Ja jos jo kansallinen tilanne on jakautunut, vielä vähemmän voi tietää kansainvälisestä tilanteesta. Osittain tämä johtuu siitä että DCMI:n ja yritysten kuten Microsoftin ja Sonyn solmimat sopimukset ovat luottamuksellisia.

Kansalliskirjasto on ottamassa entistä aktiivisempaa roolia Dublin Coren ylläpidossa Suomessa. Kirjaston sisällä formaatin tuntemusta lisätään ja jaetaan usealle henkilölle. Uusia organisaatioita on tulossa formaatin käyttäjiksi kirjastosektorilla ja sen ulkopuolella, ja erilaiset toimijat kaipaavat omia ohjeita formaatin soveltamisesta.

Kansalliskirjasto ylläpitää myös konversio-ohjelmia Dublin Coren ja muiden formaattien välillä. DCMI:n johtokunta on esittänyt mahdollisuutta järjestää kansainvälisten kongressien lisäksi myös kansallisia tapahtumia - Suomessa voisi juuri nyt olla hyvä hetki koota nykyiset ja tulevat käyttäjät yhteen keskustelemaan ja oppimaan lisää.

Dublin Coren tulevaisuus

Dublin Coren lyhyt historia on mielestäni melkoinen menestystarina, hapuilevasta alusta nykytilanteeseen, jossa Dublin Core on vakiintunut ja laajalti käytetty työkalu. Mutta jähmettyä se ei saa. Formaattiin on jatkuvasti kehitettävä uusia ominaisuuksia (termejä, profileja), jotta sen sovellettavuus paranisi entisestään. Muuttumattomuudesta tulee nopeasti metadataformaatin turma, koska sovellusympäristöt muuttuvat. Jokainen formaatti tulee vastaamaan tähän haasteeseen omalla tavallaan; niin Dublin Core kuin MARC ja sen tulevat seuraajat.

Uskon vahvasti siihen että Dublin Coren menestys jatkuu tulevaisuudessakin. Internet-yhteisö tarvitsee metadataratkaisua, jonka sisältöön se pystyy itse vaikuttamaan ja jonka kehittäminen ei ole kiinni jonkin yksittäisen toimijan hyvästä tahdosta.

Olin Dublin Core -kehitystyössä mukana noin 15 vuoden ajan, aina huhtikuussa 1996 pidettyä järjestyksessä toisesta DC-kokouksesta lähtien. Ammatillisesti nämä vuodet olivat antoisia. Jokaisen metadatan parissa töitä tekevän olisi hyvä tutustua ainakin pariin kolmeen formaattiin, koska silloin voi nähdä omankin työkalunsa toisessa valossa, ja ymmärtää paremmin semanttisen yhteismitallistamisen haasteita. Vaikka luopumiseni Dublin Core -vastuista oli vapaaehtoista ja tarpeen uusien haasteiden vuoksi, pientä haikeutta on vaikea välttää. Kauan eläköön tämä suuri pieni formaatti! 📖

Lähdeluettelo

Glossary/DCMI Abstract Model. DCMI 2011.
Saatavissa: [http://wiki.dublincore.org/index.php/Glossary/DCMI Abstract Model](http://wiki.dublincore.org/index.php/Glossary/DCMI_Abstract_Model)

Lagoze, Carl: The Warwick Framework: A Container Architecture for Diverse Sets of Metadata. D-Lib magazine, July/August 1996. Saatavissa: <http://dlib.org/dlib/july96/lagoze/07lagoze.html>

Weibel, Stu: Metadata: The Foundations of Resource Description. D-Lib magazine, July 1995. Saatavissa: <http://dlib.org/dlib/July95/07weibel.html>

Weibel, Stu & Lorcan Dempsey: The Warwick Metadata Workshop: A Framework for the Deployment of Resource Description. D-Lib magazine, July/August 1996. Saatavissa: <http://dlib.org/dlib/july96/07weibel.html>

Weibel, Stu & Eric Miller: Image Description on the Internet: A summary of the CNI/OCLC Image Metadata Workshop September 24 - 25, 1996. Dublin, Ohio. D-Lib magazine, January 1997. Saatavissa: <http://dlib.org/dlib/january97/oclc/01weibel.html>

Weibel, Stu & Renato Iannella: The 4th Dublin Core Metadata Workshop Report. March 3 -5, 1997. National Library of Australia, Canberra. D-Lib magazine, June 1997. Saatavissa: <http://dlib.org/dlib/june97/metadata/06weibel.html>

Weibel, Stu & Juha Hakala: DC-5: The Helsinki Metadata Workshop. A report on the Workshop and Subsequent Developments. D-Lib magazine, February 1998. Saatavissa: <http://dlib.org/dlib/february98/02weibel.html>

Weibel, Stu & Traugott Koch: The Dublin Core Metadata Initiative: mission, current activities and future directions. D-Lib magazine, December 2000. Saatavissa: <http://www.dlib.org/dlib/december00/weibel/12weibel.html>

Tietoa kirjoittajasta

*Juha Hakala, erityisasiantuntija
Kansalliskirjasto, Hallinto- ja kehittämispalvelut
Email. Juha.hakala@helsinki.fi*