

Dublin Core -perehdytys

Tanja Vienonen

1.4.2020, klo 13–14.30

Yleistä Dublin Coresta

- Voittoa tavoittelemattoman [Dublin Core Metadata Initiativen](#) (DCMI) ylläpitämä formaatti
 - Käytetään elektronisten/digitaalisten aineistojen kuvailuun
- Perustuu W3C:n [Research Description Frameworkiin](#) (RDF)
- Standardoitu ISO:ssa (ISO 15836-1:2017 ja ISO 15836-2:2019)
 - Suomen kansallisen Dublin Core -version uusi SFS-standardi ilmestyy tänä vuonna
- Tavoitteena aineiston haettavuuden parantaminen, ei ”täydellinen” kuvailutieto/metatieto
- Dublin Core on todennäköisesti maailman laajimmin käytetty metatietoformaatti
 - Sovelletaan laajasti kuvailun välineenä ja vaihtoformaattina
- Kirjastoalalla formaattia käytetään metatiedon haravoinnin lisäksi myös esim. julkaisuarkistoissa ja muissa sovelluksissa, joissa kuvailusta voi vastata joku muu kuin kirjaston henkilökunta

Dublin Core Kansalliskirjastossa

- Kansalliskirjastolla kansallinen ylläpito- ja kehittämisvastuu Dublin Core -formaattista
 - Sama vastuu kuin MARC 21 -formaattista
- Kansalliskirjasto DCMI:n Regional Member
 - Osma Suominen DCMI Governing Boardin ja [Usage Boardin](#) jäsen
- Yleinen asiakaswiki: <https://www.kiwi.fi/display/DublinCore>
 - Erityisesti ohjeet-sivu: <https://www.kiwi.fi/display/DublinCore/Ohjeet>
- Metadatasuositus julkaisuarkistojen tekstiaineistolle:
<https://www.kiwi.fi/x/94R7B>

Dublin Core -formaatin eri versiot

1. Dublin Core Metadata Element Set

- 15 ydinominaisuutta (elementtiä), sisältyvät DCMI Metadata Terms -nimialueeseen

2. Dublin Core Qualifiers (Qualified Dublin Core)

- Ydinelementtien ”laajennusosa” v. 1997, korvattu myöhemmin DCMI Metadata Terms -nimialueella
- Edelleen laajasti käytössä

3. DCMI Metadata Terms

- Uusin suositus 2008, päivitetty 2019
- 55 ominaisuutta (property) ja 20 luokkaa (class)

Dublin Core Metadata Element Set

- Tunnetaan myös nimellä Simple/Simplified Dublin Core
- Elementtien/ominaisuuksien määrä rajattu pieneksi käytön helpottamiseksi
- 15 ydinominaisuutta/elementtiä:
 - Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights
- Vain suppea kuvailu mahdollista
 - Käytännössä aineiston kuvailussa tarvitaan melkein aina lisäkenttiä
- Sovelletaan usein vaihtoformaattina, kun metatietoa siirretään järjestelmästä toiseen
 - Esim. OAI-PMH -rajapinta käyttää oletusarvoisesti tätä Dublin Coren versiota

Dublin Core Qualifiers 1/2

- Tunnetaan myös nimellä Qualified Dublin Core (QDC)
- DCMI Metadata Terms on korvannut tarkenteiden käytön suosituksena, mutta QDC on edelleen laajasti käytössä monissa järjestelmissä, kuten julkaisuarkistoissa
- Ydinelementteihin lisätty kolme uutta elementtiä ja tarkenteet, joilla voi rajata elementin sisältöä
 - Lisäkentät: Audience, Provenance ja Rights Holder
 - Tarkenteita voi käytännössä lisätä loputtomasti

Dublin Core Qualifiers 2/2

- Tarkenteita ei voi käyttää yksinään, ne pitää sitoa johonkin elementtiin
 - Esim. dc.relation.isversionof
 - Elementteillä vaihteleva määrä tarkenteita
 - Esim. Relation-elementin tarkenteet: Is Version Of, Has Version, Is Replaced By, Replaces, Is Required By, Requires, Is Part Of, Has Part, Is Referenced By, References, Is Format Of, Has Format, Conforms To
- DCMI:n hyväksymien tarkenteiden lisäksi voidaan luoda omia, enemmän tai vähemmän virallisia tarkenteita
 - Esim. **Metadatasuosituksessa julkaisuarkistojen tekstiaineistolle** on kansallisia lisäyksiä, kuten dc.subject.yso

DCMI Metadata Terms 1/3

- Tunnetaan myös nimellä DCMI Terms
- Tuorein Dublin Core -versio, otettu huomioon linkitetyn datan vaatimukset
- Sisältää molemmat aiemmat Dublin Core -versiot
 - Qualified Dublin Coren tarkenteet ovat itsenäisiä ominaisuuksia, myös joitain uusia ominaisuuksia lisätty
 - 15 ydinominaisuutta (element), 40 ominaisuutta (property) ja 20 luokkaa (class)
 - Nimialueen osat (kaikki 75) kuuluvat ISO 15836-2 -standardiin (muut DCMI:n sivuilla olevat tiedot ovat näiden kenttien arvoja tai käytettäviä koodaustapoja, oletan)

DCMI Metadata Terms 2/3

- Jako luokkiin ja ominaisuuksiin perustuu [Resource Description Framework Schemaan](#)
 - DCMI Terms -nimialueella luokat ovat ominaisuusjoukkoja (?)
 - Esim. Agent-luokalla on ominaisuudet Creator, Contributor ja Publisher, joita käytetään varsinaisessa kuvailussa
 - *Onko RDF:n mukaan OK käyttää kuvailussa myös luokkia, vai onko tarkoitus käyttää vain ominaisuuksia?*
- Näyttää siltä, ettei DCMI:n sivuilla olevissa sovellusprofiileissa ole käytetty luokkia (profiilien linkit myöhemmässä diassa)
- [RDF](#):n soveltamiseen liittyy myös 15 ydinelementin sisällyttäminen DCMI Terms -nimialueeseen, jossa niiden käyttötapa on tarkemmin määritelty kuin Dublin Core Metadata Element Set -nimialueella

DCMI Metadata Terms 3/3

- Formaatin laajentaminen ei ole yhtä helppoa kuin QDC:n, olemassa olevien kenttien pitäisi riittää
- Toistaiseksi vaikea käyttää e-julkaisujen metatietojen tallentamiseen, koska tarpeellisia kenttiä puuttuu
 - Ominaisuuksien valinnassa käytettyjä sovellusprofiileja: DC-Education ja DC-Collections
 - DCMI Usage Boardille lähetetään lähiaikoina formaatin laajennusehdotus

Sovellusprofiilit (Application Profiles) 1/2

- Erilaisia erikseen kehitettyjä sovellusprofiileja (Application Profiles)
 - Osa pohjaa edelleen QDC-formaattiin
 - Dublin Core -formaatin soveltamisohjeita erilaisia käyttötarkoituksia varten
 - Kehitetty vastaamaan eri yhteisöjen tarpeita
 - Virallisia, DCMI-yhteisön laatimia ja hyväksymiä profiileja, sekä muiden laatimia ratkaisuja; jälkimmäiset eivät ole ”virallisia” ja niissä voi olla kenttiä useista eri formaateista
- Esimerkkejä:
 - [SWAP](#): Scholarly Works Application Profile
 - [Metadatasuositus](#): kansallinen sovellusprofiili julkaisuarkistojen tekstiaineiston kuvailuun

Sovellusprofiilit 2/2

- DCMI on päivittänyt laatimansa tai hyväksymänsä profiilit DCMI Terms -nimialueen mukaisiksi
- Päivitettyjä sovellusprofiileja (QDC ei enää käytössä):
 - [Dublin Core Collection Description](#): Collections Application Profile
 - [DC-Libraries](#): Library Application Profile
 - [DCAT](#): Data Catalog Vocabulary (W3C:n laatima, useampi formaatti käytössä)
- [DCMI Application profiles Interest Group \(IG\)](#)
 - [Singapore Framework for Dublin Core Application Profiles](#)

Metadatasuositus julkaisuarkistojen tekstiaineistolle 1/4

- Kansallinen sovellusprofiili QDC-formaatin pohjalta
- Tavoitteena koota mahdollisimman käyttökelpoinen, tarpeeksi laaja ja monipuolinen sovellusprofiili
 - Metatieto tarpeeksi rikasta ja tarkasti eroteltua, jotta ulospäin välitettävät formaatit voidaan tuottaa
- Pyritty löytämään kaikille kuvailutiedoille oma, selkeä kenttänsä ja ohjeistaa niiden käyttö yksiselitteisesti
 - Pohjana julkaisuarkistoissa jo käytössä olleet kentät
 - Kenttien koneluettavuuteen pyritty
- Metatiedon loogisuus ja johdonmukaisuus
 - Parempi vähän ja paikkansapitävää metatietoa, kuin ei ollenkaan tai virheellistä metatietoa

Metadatasuositus julkaisuarkistojen tekstiaineistolle 2/4

- Kenttiä alun perin 62, päivityksen jälkeen 77 – QDC:ssä kenttiä 51 ≈ 26 kansallista kenttää
 - Kaikkia QDC:n kenttiä ei käytetä, esim. Audience, RightsHolder ja osa Date-kentän tarkenteista
 - Osaa kentistä käytetään standardinvastaisesti järjestelmän vaatimusten vuoksi
 - Esim. Provenance (muokkaushistoria, ei omistushistoria), dc.format.extent (sivumäärä, ei tiedoston koko)
- Joissain kentissä annettu ”lupa” laajentaa suositusta edelleen
 - Esim. dc.contributor.[muu tekijyys], dc.identifier.[tunnisteen lyhenne], dc.relation.[tunnisteen lyhenne], dc.subject.[asiasanaston lyhenne]
 - Kentissä annettu suositus käytettävistä tarkenteista ja tallennettavista arvoista

Metadatasuositus julkaisuarkistojen tekstiaineistolle 3/4

- Kansallisia kenttiä:
 - OKM:n tiedonkeruuseen liittyvät kentät ns. omana nimialueenaan
 - Ei ole tarkoitus, että haravoituu muualle, informaatioarvo pieni muille kuin OKM:lle
 - Kaikki dc.contributor-, dc.identifier- ja dc.rights-kentän tarkenteet ovat kansallisia
 - Puolet dc.relation-kentän tarkenteista on kansallisia (jotkut vahingossa...)
- Suosituksen mukaisia syöttölomakkeita on tehty suurimmalle osalle KK:n ylläpitämistä julkaisuarkistoista
 - Takautuvaa metatietoa on päivitetty niin pitkälle kuin on pystytty niillä resursseilla mitä organisaatioilla on
 - Päivitystarve tulee asiakkailta, mutta usein kaupataan järjestelmäpäivitysten yhteydessä
 - Organisaatiot ovat vastuussa julkaisuarkistojen sisällöstä ja sitä myöten myös siellä olevasta metatiedosta
 - Osa on rajannut metatiedon päivityksen vain uudempiin, yhä karttuviin aineistoihin

Metadatasuositus julkaisuarkistojen tekstiaineistolle 4/4

- Suositus on otettu soveltuvin osin käyttöön ainakin näissä julkaisuarkistoissa:
 - Ammattikorkeakoulujen Theseus
 - Jyväskylän yliopiston JYX
 - LUT-yliopiston LutPub
 - Suomen Ympäristökeskuksen Jukuri (uudet aineistot)
 - STM:n hallinnonalan avoin julkaisuarkisto Julkari (ainakin uudet aineistot)
 - Tampereen yliopiston Trepo (ainakin aiemman Tampereen yliopiston aineistojen osalta)
 - Turun yliopiston UTUPub
 - Vaasan yliopiston Osuva
- KK:n ylläpitämistä päivittämättä:
 - Doria, 11 yhä karttuvaa yhteisöä
 - Fenno-Ugrica
 - Fragmenta Membranea
 - Lapin yliopiston Lauda
 - Valtioneuvoston Valto (ehkä)

Kysyjälle vastataan

- Julkaisuarkistot ja kk-, qdc_finna- ja qdc_kk-haravointiformaatit: doria-oa@helsinki.fi ja Samu Viita
- Dublin Core ylipäätään: dc-posti@helsinki.fi, Tanja Vienonen ja Osma Suominen
- Standardit: Juha Hakala

Kiitos!