

Dublin Core ja sen käyttö WWW-sivuilla

Timo Aittokoski, Miika Nurminen, Sami Äyrämö

21. huhtikuuta 2004

Tiivistelmä

Metadata (tai metatieto) on yleisesti sanottuna ”tietoa tiedosta”. Metadataa käytetään, paitsi kuvailemaan ja identifioimaan tietoa eli informaatio-objekteja, niin myös dokumentoimaan niiden käyttäytymistä, toimintaa, käyttöä ja suhdetta muihin informaatio-objekteihin sekä sitä, kuinka niiden sisältämää tietoa tulisi käsitellä. Nykyaikana metadatan roolia korostetaan erityisesti kehitettäessä tehokkaita, asiantuntevia, skaalautuvia ja ylläpidettäviä tietojärjestelmiä. Metadataa on hyödynnetty mm. informaatio-objektien sisällön autenttisuuden ja täydellisyyden todentamiseen, kontekstin määrittämiseen ja dokumentointiin tai jonkin muun lisäinformaation tarjoamiseen. Uudemmissa alueilla metadataa voidaan hyödyntää esimerkiksi Internet-sivustojen etsinnässä HTML-komentojen avulla.

Dublin Coren elementtejä käytetään verkossa saatavilla olevien resurssien (yleensä dokumenttien) kuvailemiseen ja identifioimiseen. Elementtien metadata on suositeltavaa luoda tiettyjen suositusten mukaan. Elementin arvo voidaan tietyissä elementeissä valita ”kontrolloiduista sanastoista”, jotka ovat rajallinen kokoelma huolellisesti määriteltyjä ja johdonmukaisesti käytettyjä termejä. Elementit voidaan karkeasti luokitella kolmeen eri ryhmään siten, että osa elementeistä liittyy kohteen sisältöön, osa tekijätietoihin ja osa kohteen tunnistamiseen (versioon).

Dublin Core -standardi tasapainoilee kahden asian kanssa: Dublin Core on pidettävä yksinkertaisena mutta sen oltava myös tehokas ja tarkka apuväline tiedonhaussa. Tästä syystä Dublin Coren elementtejä voidaan laajentaa, jotta se palvelisi paremmin vaativia tiedontarpeita. Tarkenteiden (*qualifier*) käyttäminen antaa eri tahoille mahdollisuuden käyttää Dublin Core-kuvailua tallenteiden ”ytimen” kuvailuun mutta samalla se mahdollistaa myös tarkemman kuvailun spesifeihin tarpeisiin.

Metatieto voidaan sijoittaa suoraan sivujen sisälle tai erillisiin kuvailutiedostoihin. Dublin Coren esitystapoja ovat HTML-kielen <meta>- ja <link>-elementit, XML-kielinen esitystapa omassa nimiavaruudessaan tai RDF-kuvaus. Näistä RDF on ilmaisuvoimaisin mahdollistaen eri metatieto-standardien yhdistämisen, tietorakenteet ja väitteiden ilmaisemisen toisista väitteistä.

Harjoitustyön työnjako on toteutettiin seuraavasti: Sami Äyrämö vastasi metadataa käsittelevästä luvusta, Timo Aittokoski Dublin Coren elementtien kuvauksesta ja Miika Nurminen WWW-esimerkeistä.

1 Metadata

Tässä kappaleessa kuvataan yleisesti, lähteisiin [13] ja [14] perustuen, mitä on metadata, mihin sitä käytetään ja mitä ominaisuuksia sillä on.

1.1 Mitä on metadata?

Metadatan (tai metatieto) voidaan sanoa olevan ”tietoa tiedosta” (engl. *data about data*). Metadatalta on hieman toisistaan eroavia merkityksiä eri ammattipiireissä, jotka suunnittelevat, luovat, kuvaavat, ylläpitävät ja käyttävät tietojärjestelmiä ja niihin liittyviä resursseja. Nykyaikana metadatan roolia korostetaan erityisesti kehitettäessä tehokkaita, asiantuntevia, skaalautuvia ja ylläpidettäviä tietojärjestelmiä.

Sivistysperinnön ja tietopalvelun ammattilaisten keskuudessa, kuten museorekistereiden, kirjastoluetteloiden ja arkistojen ylläpitäjät, metadatalta on perinteisesti tarkoitettu sellaista tietoa, jota hyödynnetään aineistoja järjestellessä, luetteloidessa, kuvailtaessa tai muulla tavoin sen saatavuutta parannettaessa. Kirjastot ovat olleet metadatan kehittämisessä edelläkävijöitä, koska tarvittavat sisällöt pitää voida etsiä mahdollisimman tehokkaasti. Kirjastometadatta ovat esimerkiksi indeksit ja abstraktit.

Laajemmin katsottuna Gilliland-Swetland [13] toteaa metadatatista seuraavasti:

The sum total of what one can say about any information object at any level of aggregation.

Informaatio-objektilla (engl. *information object*) Gilliland-Swetland tarkoittaa tässä yhteydessä mitä tahansa asiaa, johon voidaan viitata tai jota voidaan manipuloida erillisenä kokonaisuutena ihmisen tai koneen toimesta. Informaatio-objekti voi olla yksi objekti tai se voi olla myös kooste useammasta pienestä yksittäisestä asiasta. Riippumatta informaatio-objektin tyylistä, siihen voidaan liittää aina seuraavat kolme piirrettä, joihin metadatan avulla voidaan viitata:

- *Sisältö* viittaa siihen mitä objekti on, tai mitä se sisältää. Sisältö määräytyy objektin sisäisistä ominaisuuksista.
- *Konteksti* määrittää kuinka erilaiset aspektit, kuten *kuka, mikä, miksi, missä tai kuinka*, liittyvät objektin luomiseen. Konteksti määräytyy objektin ulkopuolelta.
- *Rakenne* liittyy systemaattiseen joukkoon assosiaatioita, jotka voivat olla joko yksittäisen objektin sisällä tai useamman itsenäisen objektin välillä. Rakenne voi olla ulkoinen tai sisäinen.

1.2 Metadatan käyttö

Sellaisissa tietoverkoissa, joissa käyttäjän pääsee suoraan tietoverkon kautta käsiksi informaatio-objekteihin, metadatta voidaan käyttää:

- sisällön autenttisuuden ja täydellisyyden todentamiseen.
- sisällön kontekstin määrittämiseen ja dokumentointiin.
- informaatio-objektien sisäisten ja niiden välillä olevien rakenteisten suhteiden tunnistamiseen ja hyödyntämiseen.

- tarjoamaan älykäs ja luokiteltu pääsy informaatioon käyttäjälukille.
- tarjoamaan lisäinformaatiota, jota informaatioalan ammattilainen on hankkinut esimerkiksi tutkimusympäristöstään.

Uudemmissa alueilla, kuten Internetissä, palveluntarjoaja voi käyttää HTML-komentojen avulla metadataa helpottaakseen sivustonsa löytämistä. Digitaalisten kuvien käsittelijä voi käyttää metadataa lisätessään kuvatiedoston otsikkoon tietoa kuvasta ja siihen liittyvistä tekijänoikeuksista. Metadataa voi käyttää muuhunkin kuin tiedon kuvailemiseen ja identifiointiin. Sen avulla voidaan dokumentoida myös objektin käyttäytymistä, toimintaa ja käyttöä, objektin suhdetta muihin objekteihin sekä sitä, kuinka tietoa tulisi käsitellä.

1.3 Metadatan luokittelu

Yleisesti ymmärrys siitä, mitä metadatalta tarkoitetaan on hyvin laeva. Kaikki eri näkökulmat metadataan ovat tärkeitä kehitettäessä digitaalisia tietoverkkojärjestelmiä. Jotta metadatan käsitteenä olisi yksinkertaisempi ymmärtää, se on jaettu viiteen erilliseen luokkaan: *hallinnollinen*, *kuvaileva*, *tekninen*, *käyttö* ja *säilyttäminen*. Luokkien tarkemmat kuvaukset on esitetty taulukossa 1.

Tyyppi	Määritelmä	Esimerkkejä
Hallinnollinen	Tietoresurssien hallinnassa ja hoidossa käytetty metadatan	<ul style="list-style-type: none"> • Tietoja hankinnoista • Oikeudet ja toisintamisen seuranta • Lain edellyttämän tietosuojan dokumentointi • Tietoja sijainnista • Valintaperusteet digitalisoinnille • Versioiden kontrolli
Kuvaileva	Tietoresurssien kuvailussa tai tunnistamisessa käytetty metadatan	<ul style="list-style-type: none"> • Luetteloitirekisterit • Hakupalvelut • Erikoishakemistot • Hyperlinkit resurssien välillä • Käyttäjien huomautukset
Tekninen	Järjestelmän toimivuuteen tai metadatan käyttäytymiseen liittyvä metadatan	<ul style="list-style-type: none"> • Laitteisto- ja ohjelmistodokumentointi • Sähköistämisen tiedot, esim. formaatit, tiivistyssuhteet, skaalausruutit • Järjestelmän vasteaikojen jäljitys • Elektroninen tunnistus ja varmistus, esim. salausavaimet ja salasanat
Metadatan säilyttäminen	Tietoresurssien säilyttämisen järjestämisen liittyvä metadatan	<ul style="list-style-type: none"> • Resurssien fyysisen tilan dokumentointi • Resurssien fyysisten ja digitaalisten versioiden säilyttämiseen tarvittavien toimien dokumentointi, esim. tietojen päivitys ja muutto
Metadatan käyttö	Tietoresurssien käytön tasoa ja tyyppiä kuvaava metadatan	<ul style="list-style-type: none"> • Esitystiedot • Käytön ja käyttäjän jäljitys • Tiedot uudelleenkäytöstä ja versioista

Kuva 1: Metadatan eri luokat ja niiden toiminnot. [14]

Metadataan liittyvien luokkien ja toimintojen lisäksi metadatalta on myös monia erilaisia ominaisuuksia. Taulukossa 2 on esitetty joitakin metadataan liittyviä avainattributteja piirteineen.

Ominaisuus	Luonteenpiirre	Esimerkkejä
Metadatan lähde	<ul style="list-style-type: none"> • Sisäinen metadata, joka on määritelty tiedosta kohteen luomisen tai digitalisoinnin aikana ◆ Ulkoinen metadata, joka on luotu myöhemmin, usein jonkun muun kuin alkuperäisen luojan toimesta 	<ul style="list-style-type: none"> • Tiedostonimet a otsikkotiedot • Hakemistorakenteet • Tiedostoformaattit ja pakkausmenetelmät ◆ Rekisteri- ja luettelointitiedot ◆ Tekijän- yms. oikeudet
Metadatan luomisen metodi	<ul style="list-style-type: none"> • Tietokoneen luoma automaattinen metadata ◆ Ihmisten luoma manuaalinen metadata 	<ul style="list-style-type: none"> • Avainsanaluettelot • Käyttäjän tapahtumalokit ◆ Kuvailtavat korvikkeet kuten luettelointirekisterit ja Dublin Core -metadata
Metadatan luonne	<ul style="list-style-type: none"> • "Maallikometadata", jonka luonut henkilö ei ole kohteen tai tiedon spesialisti. Tekijänä usein informaatio-objektin ⁽¹⁾ luoja ◆ "Ekspertti-metadata". Tekijänä usein joku muu kuin informaatio-objektin luoja (esim. informaattikko) 	<ul style="list-style-type: none"> • Henkilökohtaisille Web-sivuille luodut metatagit • Henkilökohtaiset tiedostojärjestelmät ◆ Erikoistuneet aiheotsikot ◆ MARC (Machine Readable Catalogues) rekisterit ◆ Arkistoista etsimiseen tarvittavat apukeinot
Metadatan tila	<ul style="list-style-type: none"> • Pysyvä metadata, joka ei muutu sen luomisen jälkeen ◆ Dynaaminen metadata, joka voi muuttua informaatio-objektia käytettäessä tai sitä ohjailtaessa ▪ Pitkän tähtäimen metadata, joka on välttämätöntä varmistamaan, että informaatio-objekti on käytettävissä ja säilytetään ➤ Lyhyen ajan metadata, jolla on toimintoa tukeva luonne 	<ul style="list-style-type: none"> • Tietoresurssin otsikko, alkuperä ja luomispäivä ◆ Hakemistorakenne ◆ Käyttäjälokit ◆ Kuvan resoluutio ▪ Tekninen muoto ja käsittelyinformaatio ▪ Käyttöoikeustiedot ➤ Tiedon säilyttämisen dokumentointi
Metadatan rakenne	<ul style="list-style-type: none"> • Jäsennelty metadata, joka täyttää odotuksenmukaisen standardoidun tai standardoimattoman rakenteen ◆ Jäsentelemätön metadata, joka ei noudata odotuksenmukaista rakennetta 	<ul style="list-style-type: none"> • MARC • TEI ja EAD • Tietokantaformaatti ◆ Jäsentelemättömät huomautus- ja muistiinpanokentät
Metadatan semantiikka	<ul style="list-style-type: none"> • Kontrolloitu metadata, joka täyttää standardoidun sanaston tai virallisen muodon vaatimukset ◆ Kontrollioimaton metadata, joka ei täytä edellisten vaatimuksia 	<ul style="list-style-type: none"> • AAT • ULAN • AACR2 ◆ Vapaateksti -huomautukset ◆ HTML -metatagit
Metadatan taso	<ul style="list-style-type: none"> • Kokoelmatason metadata, joka liittyy kokoelmiin ◆ Yksikkötason metadata, joka liittyy yksittäisiin informaatio-objekteihin, jotka usein sisältyvät kokoelmiin 	<ul style="list-style-type: none"> • Kokoelmatason rekisterit, esim. MARC tai arkistotietokantojen hakupalvelut • Erityiset indeksit ◆ Sovitetut kuvatestit ja päivämäärät ◆ Formaatti

Kuva 2: Avainattribuutteja ja piirteitä metadatalle. [14]

2 Dublin Coren elementit

Dublin Coren elementtejä käytetään verkossa saatavilla olevien resurssien (yleensä dokumenttien) kuvailemiseen ja identifioimiseen. Eri elementtejä on 15 kappaletta ja ne koostuvat (*nimi, arvo*) -pareista, esimerkiksi: Title="The Sound of Music"

Jokainen elementti on vapaaehtoinen ja toistettavissa. Metadata-elementtien järjestyksellä ei ole merkitystä. Useita samoja kenttiä sisältävässä kuvailussa on kuvailun tekijällä saattanut olla jokin järjestysperustelu, mutta ei ole taattu, että se pysyisi samana eri käyttäjäympäristöissä.

Elementtien metadata on suositeltavaa luoda tiettyjen suositusten mukaan. Eräs tällainen suositus on Dublin Core -organisaation sivuilla¹. Usein kuvailutieto voidaan sijoittaa selkeästi tiettyyn elementtiin, mutta toisinaan joidenkin elementtien välillä saattaa olla semanttista päällekkäisyyttä. Tämän vuoksi kuvailutiedon lisääjän täytyy käyttää omaa arviointikykyään.

Elementin arvo voidaan tietyissä elementeissä valita ”kontrolloiduista sanastoista”, jotka ovat rajallinen kokoelma huolellisesti määriteltyjä ja johdonmukaisesti käytettyjä termejä. Näiden käyttö voi parantaa huomattavasti hakutuloksia, koska tietokoneet ovat hyviä vertailemaan sanoja merkki merkiltä mutta huonoja ymmärtämään ihmisten käyttämiä synonyymejä eri käsitteille. Esimerkiksi ilman kontrolloitua sanastoa. Termejä ”karamellit”, ”karkit” ja ”makeiset” voidaan käyttää tarkoittamaan samaa asiaa (käsitettä). Sanastojen käyttö saattaa myös vähentää kirjoitusvirheiden määrää metadatan tallennusvaiheessa. Suomessa kuvailuun ja tiedonhakuun voidaan käyttää Yleistä Suomalaista Asiasanastoa, jota ylläpitää Helsingin yliopiston kirjasto.

Elementit voidaan karkeasti luokitella kolmeen eri ryhmään siten, että osa elementeistä liittyy kohteen sisältöön, osa tekijätietoihin ja osa kohteen tunnistamiseen (version).

- Sisältöön liittyvät elementit:
 - Coverage
 - Description
 - Type
 - Relation
 - Source
 - Subject
 - Title
 - Audience
- Tekijätietoihin liittyvät elementit:
 - Contributor
 - Creator
 - Publisher
 - Rights
- Versioon liittyvät elementit:

¹<http://dublincore.org/documents/usageguide/elements.shtml>

- Date
- Format
- Identifier
- Language

Dublin Core -standardi tasapainoilee kahden asian kanssa: Dublin Core on pidettävä yksinkertaisena mutta sen oltava myös tehokas ja tarkka apuväline tiedonhaussa. Tästä syystä Dublin Coren kehittäjät ovat pitäneet tärkeänä tarjota mekanisme, jonka avulla Dublin Core elementtejä voidaan laajentaa, jotta se palvelisi paremmin vaativia tiedontarpeita. Tämä malli antaa eri tahoille mahdollisuuden käyttää Dublin Core-kuvailua tallenteiden ”ytimen” kuvailuun mutta samalla se mahdollistaa myös tarkemman kuvailun spesifeihin tarpeisiin. Tarkenne (*qualifier*) on metadataformaatis- sa käytetty keino tarkempien spesifien kuvausten tekemiseen. Tarkenteet jaetaan kenttätarkenteisiin ja merkintäjärjestelmiin. Kenttätarkenne (*element qualifier*) on tarkenne, joka muokkaa kentän merkitystä yksityiskohtaisem- maksi Merkintäjärjestelmä (*encoding scheme*) on tarkenne, jolla viitataan jonkin standardin tai konvention mukaiseen merkintätapaan. Dublin Core - yhteisö hyväksyi tarkenteet kesällä 2000. 15 peruskentän lisäksi formaatin soveltaja voi ottaa käyttöön omia kenttiä ja tarkenteita. [1] [2] [4] [12]

Elementti	Tarkenne
Title (Nimeke) Tekijän, laatijan tai julkaisijan antama asiakirjan nimi, otsikko tai asia.	Alternative
Creator (Tekijä) Henkilö tai yhteisö, joka on ensisijaisesti vastuussa tallenteen intellektuaalisesta sisällöstä. Esim. kirjoittaja (kirjoitetuille dokumenteille), taiteilija, valokuvaaja tai kuvittaja (kuvadokumenteille).	-
Subject (Aihe) Tallenteen aihealueen kuvaus, yleensä asiasanoilla tai fraaseilla kuvattuna. Kontrolloitujen sanastojen ja luokitusten käyttö on suositeltavaa.	-
Description (Kuvaus) Tallenteen sisällön vapaamuotoinen kuvaus. Esimerkiksi tekstidokumentin tiivistelmä tai visuaalisen tallenteen sisällönkuvaus.	Table Of Contents Abstract
Publisher (Julkaisija) Organisaatio, joka on julkaissut tai julkistanut tallenteen sen nykyisessä muodossa. Esimerkiksi kustannusyhtiö, yliopiston laitos tai jokin muu yhteisö.	-

<p>Contributor (Muu tekijä) Henkilö tai organisaatio, joka Tekijä-kentässä mainittujen henkilöiden / organisaatioiden lisäksi on osallistunut merkittävästi tallenteen luomiseen, mutta jonka panos on kuitenkin ollut toissijainen verrattuna ns. päävastuullisen tekijän osuuteen. Esimerkiksi toimittaja, kääntäjä tai kuvittaja.</p>	-
<p>Date (Päivämäärä) Päivämäärä, jolloin tallenne on julkaistu tai julkistettu kyseisenä versiona. Tätä päivämäärää ei pidä sekoittaa Kate-kentässä esiintyviin aikamerkintöihin. Suositeltava tallennusmuoto on VVVV-KK-PP, kuten on mritelty ISO 8601:n profiilissa². Esimerkiksi 1998-08-03 tarkoittaa elokuun kolmatta 1998.</p>	<p>Created Valid Available Issued Modified Date Copyrighted Date Submitted</p>
<p>Type (Laji) Tallenteen laji, esimerkiksi kotisivu, romaani, runo, artikkeli, sanakirja tai valokuva. Kuvailun yhteismitallisuuden takia laji tulisi valita Dublin Core -yhteisön yllpitämästä listasta, joka ei ole yhteismitallinen FINMARC-formaatin kirjallisuuslaji- ja teoksen sisältökoodien kanssa.</p>	-
<p>Format (Formaatti) Tallenteen tiedostformaatti, kuten teksti/html, sovellus tai JPEG-kuva. Formaattitietoa voidaan käyttää sen selvittämiseen, mitä laitteisto- ja ohjelmistoympäristöä tallenteen hyödyntäminen vaatii.</p>	<p>Extent Medium</p>
<p>Identifier (Identifikaatiotunnus) Tunnus (merkkijono tai numero), joka yksiselitteisesti identifioi tallenteen. Esimerkiksi verkkodokumenttien URN tai URL-tunnukset (URL:n eli sijaintitiedon käyttöä ainoana tunnuksena ei suositella).</p>	<p>Bibliographic Citation</p>
<p>Source (Lähde) Tunnus (merkkijono tai numero), joka yksiselitteisesti identifioi tallenteen lähteen, jos sellainen on olemassa.</p>	-
<p>Language (Kieli) Tallenteen kieli. Kieli tallennetaan ISO 639-1 -standardin mukaisena kaksikirjaimisena koodina (esim. en, fi, fr).</p>	-

²<http://www.w3.org/TR/NOTE-datetime>

<p>Relation (Suhde)</p> <p>Ilmaistaan toisen tallenteen identifikaatiotunnus, joka on jossakin suhteessa kuvailtavaan tallenteeseen. Elementissä voidaan myös ilmaista kuvailtavan tallenteen suhde muihin tallenteisiin (ks. suhdetyypit alla). Esimerkiksi dokumentin sisältämät kuvat, kirjan luvut tai kokoelman osat.</p>	<p>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</p>
<p>Coverage (Kate)</p> <p>Tallenteen ajallinen tai maantieteellinen kate. Maantieteellinen kate viittaa fyysiseen alueeseen; koordinaatteihin tai paikannimeen. Ajallinen kate viittaa lähinnä tallenteen asiasisltöön, ei siihen koska se on luotu tai asetettu saataville (kuuluvat Päivämäärä-elementtiin).</p>	<p>Spatial Temporal</p>
<p>Rights (Tekijänoikeudet)</p> <p>Lyhyt maininta tallenteen käyttöoikeuksista tai linkki tallenteen tekijänoikeustietoihin tai palveluun, joka antaa tietoja tallenteen käytöstä ja oikeuksista.</p>	<p>Access Rights</p>
<p>Audience (Kohderyhmä)³</p> <p>Ryhmä, jolle tallenne on suunnattu tai katsottu hyödylliseksi. Kohderyhmän määrittää Tekijä, Julkaisija tai muu kolmas osapuoli.</p>	<p>Mediator Education Level</p>

Taulukko 1: Dublin Coren elementit ja tarkenteet

3 Dublin Coren käyttö WWW-sivuilla

WWW:n kannalta metatiedolla kuvataan resursseja. Resurssi tarkoittaa tässä mitä tahansa WWW:n kautta saatavissa olevaa, esim. (X)HTML-sivua tai sen osaa, kuvaa tai sähköpostiosoitetta. Resurssi voi kuvata myös tunnistettavissa olevaa todellisen maailman oliota, jolle ei ole vastinetta WWW:ssä, esim. taideteosta. Resurssiin viitataan URI:lla. Metatieto voidaan sijoittaa suoraan web-sivujen sisälle tai erillisiin kuvailutiedostoihin. Esitystapoja on useita.

Dublin Core -metatieto on tekstimuotoista ja siten ihmisen luettavissa, joten sitä on HTML-sivujen tapaan mahdollista tuottaa suoraan käsin. Suositeltavampaa on kuitenkin käyttää jotain työkalua, joka generoi metatietokuvaukset käyttäjän syötteistä. Helsingin yliopiston kirjastolla on käytössä ilmainen Dublin Core -tallennusala⁴, jolla voi tuottaa metatietokuvauksia helposti. Palvelun hyvänä puolena mainittakoon linkit moniin eri kontrolloituihin sanastoihin. Lisäksi palvelu tukee tarkentimia.

³Huom! Tämä ei ole osa Simple Dublin Coren viidestätoista elementistä, ja tätä pitäisi käyttää vain jos käytössä ovat Dublin Coren tarkentimet.

⁴<http://www.lib.helsinki.fi/cgi-bin/dc.pl>

Dublin Core -organisaation sivuilla⁵ on laajahko lista työkaluista metadatan käsittelyyn. Varsinaisten tuotanto-ohjelmien lisäksi tarjolla on selaimia, muunnosohjelmia, automaattisia generaattoreita ja metadata-arkistojen hallintaohjelmia. Osa ohjelmistoista on kaupallisia.

3.1 Dublin Coren esitystavat

Dublin Core -metatietoa voidaan esittää suoraan HTML-sivun sisällä `<meta>`- ja `<link>`-elementeissä, minkä tahansa XML-dokumentin osana omassa nimiavaruudessaan tai RDF-määrittymisenä. RDF-määrittymis voi olla itse dokumentissa tai sen ulkopuolella. Näistä ensimmäinen tapa on perinteisin, mutta XML-sisällön lisääntymisen ja semanttisen webin kehittymisen myötä RDF-kuvausten määrä yleistynee.

3.1.1 HTML-esitys

HTML-dokumenttiin liittyvä metatieto koostuu (*ominaisuus,arvo*)-pareista. Ominaisuuksien arvot voidaan liittää dokumenttiin kahdella tavalla: [11]

1. Suoraan dokumentista `<meta>`-elementillä.
2. Viittauksena dokumentin ulkopuolelle `<link>`-elementillä.

Lisäksi `<head>`-elementtiin voidaan määrittellä `profile`-attribuutti, jossa ominaisuudet ja niiden sallitut arvot määritellään. HTML-määrittymis ei ota kantaa `<meta>`-elementtien sisältöön.

Dublin Core -ominaisuuksissa käytetään yleensä DC-etuliitettä (voidaan määrittää `<link>`-elementin avulla) erottamaan ne muista metatietomäärittymisistä. `<meta>`-elementeillä voidaan kuvata ne ominaisuudet, joiden arvo on merkkijonotietoa. `<link>`-elementillä voidaan viitata muihin dokumentteihin URI:lla. Tarkentimet ilmaistaan DCTERMS-etuliitteellä nimettyinä `<meta>`-elementteinä tai `scheme`-attribuuttina. Esimerkiksi: [9]

```
<meta name="DC.date"           content="2004-04-20" />
<link rel="DC.relation"       content="http://www.jyu.fi/" />
<meta name="DCTERMS.audience" content="software developers" />
<meta name="DC.type"         content="Text" scheme="DCTERMS.DCMIType" />
```

`<meta>`-elementeillä Dublin Core -määrittymisten kuvaaminen on yksinkertaista. Elementtien ongelmana on niiden väärinkäyttö monilla olemassaolevilla sivustoilla, jotka pyrkivät nostamaan sijoituksiaan hakukoneiden tuloslistoissa. Tämän seurauksena jotkut hakukoneet saattavat hylätä `<meta>`-elementeissä olevat tiedot kokonaan [5]. Lisäksi `<meta>`-elementit on sijoitettava kuvattavan tiedoston sisälle.

3.1.2 XML-esitys

Dublin Core -metatietoa voidaan liittää mihin tahansa XML-dokumenttiin XML-nimiavaruuksien avulla. Nimiavaruudet mahdollistavat XML-elementtien yhdistämisen samaan dokumenttiin eri skeemoista. Jokaista skeemaa vastaa käyttäjän määrittelemä nimiavaruus, jotka yksilöidään

⁵<http://dublincore.org/tools/>

URI-tunnisteilla. XML-koodissa nimiavaruudet näkyvät elementeissä etuliitteinä, jotka erotetaan varsinaisesta elementin nimestä kaksoispisteellä. [3]

Dublin Core -elementit määritellään nimiavaruudessa, jonka URI on <http://purl.org/dc/elements/1.1/>. Ominaisuuksien arvot suositellaan kirjoittamaan elementtien sisällöksi, ei attribuutteihin. Tarkennetut määrittelyt voidaan ilmaista XML-skeemaan kuuluvan `xsi:type`-attribuutin avulla: [10]

```
<dc:title>Dublin Core in XML</dc:title>
<dc:identifier xsi:type="dcterms:URI">
  http://www.ukoln.ac.uk
</dc:identifier>
```

XML-muoto on HTML-muotoa joustavampi esitys, mutta jäänee RDF-kielen varjoon. XML-muodon ongelmana on, että sovelluksen täytyy päätellä, mitkä elementit käsitellään metatietona ja mitkä sisältönä, kun taas RDF on suunniteltu erityisesti metatietoa silmälläpitäen. Lisäksi metatieto on sijoitettava suoraan kuvattavaan tiedostoon.

3.1.3 RDF-esitys

W3C:n kehittämä RDF-kieli on XML-pohjainen yleiskäyttöinen malli resurssien kuvaamiseen. RDF-kuvaus on yleistys WWW-sivuilla käytetyille `<meta>`-elementeille. RDF mahdollistaa `<meta>`-elementtejä monipuolisemman tiedon liittämisen resursseihin - jopa niin, että tiedon kuvaamisessa käytettävät käsitteet voivat olla eri lähteistä.

RDF-kuvaukset ovat kolmikkoja (*resurssi, ominaisuus, arvo*), joista jokainen voi edelleen olla resurssi. Kielitieteen käsittein metakuvaukset ovat lauseita, joissa resurssi on subjekti, ominaisuus predikaatti ja arvo objekti. Predikaattilogiikassa kolmikko vastaa kaksipaikkaista predikaattia. Resurssikuvausten joukko voidaan tulkita myös suunnattuna, tyytittynä graafina (poiketen XML:stä, jonka tietomalli on oleellisesti puumainen).

RDF-kielessä on oma linkitustoimintonsa, jolla kuvaukset liitetään resursseihin. Kielessä on vakiona yksinkertaisia säiliöluokkia, kuten jonot ja laukut. Käyttäjät voivat myös määritellä omia rakenteisia tietotyyppisiä RDFS-skeemakielen avulla. Varsinaisten resurssikuvausten lisäksi RDF tulee konkretisointia (*reification*), väitteiden tekemistä muista RDF-väitteistä. RDF-kieltä varten on määritelty kaksi XML-syntaksia (tavallinen ja lyhennetty) sekä erityinen N-triples -notaatio, joka soveltuu XML-syntaksia paremmin ihmisen luettavaksi. Merkittävää RDF-kielessä ei ole kuitenkaan syntaksi, vaan tietomalli, joka mahdollistaa monimuotoiset metakuvaukset. [7]

Dublin Core -elementtijoukon määrittelyt voidaan esittää RDFS-skeemana tai DTD:nä, jolloin metakuvauksen rakenne (mutta ei sisältöä) voidaan validoida automaattisesti. Dublin Core -kuvausten käyttö RDF-kuvauksessa edellyttää oman XML-nimiavaruuden varaamista kummallekin (esim `rdf` ja `dc`). Jokainen RDF-kuvaus esittää väitteitä yhdestä resurssista. Jokaiseen kuvaukseen voidaan upottaa useampia arvoja, jotka tässä siis ovat Dublin Core -kuvauksia.

Dublin Coren standardversion RDF-malli tukee vain tekstimuotoisia ominaisuuskuvauksia. Ottamalla huomioon tarkentimet saadaan huomattavasti ilmaisuvoimaisempi, mutta myös monimutkaisempi malli, joka

käyttää RDF-kielen kehittyneempiä ominaisuuksia hyödyksi. Tarkentimilla voidaan kuvata esimerkiksi resurssin kirjoittajien järjestys jonorakenteella. RDF:n konkretisoinnin avulla metakuvauksista voidaan edelleen tehdä “metametakuvauksia”:

```
<rdf:Description rdf:about="http://www.potato.org/">
  <dcterms:abstract rdf:ID="19">
    An excellent introduction to potato theory is
    provided by this highly original paper
  </dcterms:abstract>
  <dc:creator>John Doe</dc:creator>
</rdf:Description>
<rdf:Description rdf:about="#19">
  <dc:creator>John Doe</dc:creator>
</rdf:Description>
```

Esimerkissä on kuvaus *John Doen* kirjoittamasta artikkelista. Kuvauksen tekijäksi on edelleen merkitty konkretisoinnin avulla *John Doe*. [6]

3.2 Käyttöesimerkkejä

Käydään läpi laajemmat esimerkit kustakin Dublin Coren esitystavasta. Alla HTML-esitys DMCI:n HTML-ohjeiden [9] metatiedosta:

```
<head profile="http://dublincore.org/documents/dcq-html/">
  <title>Expressing Dublin Core in HTML/XHTML meta and link elements</title>
  <link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
  <link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
  <meta name="DC.title" lang="en"
    content="Expressing Dublin Core in HTML/XHTML meta and link elements" />
  <meta name="DC.creator" content="Andy Powell, UKOLN, University of Bath" />
  <meta name="DCTERMS.issued" scheme="DCTERMS.W3CDTF" content="2003-11-01" />
  <meta name="DC.identifier" scheme="DCTERMS.URI"
    content="http://dublincore.org/documents/dcq-html/" />
  <link rel="DCTERMS.replaces" hreflang="en"
    href="http://dublincore.org/documents/2000/08/15/dcq-html/" />
  <meta name="DC.format" scheme="DCTERMS.IMT" content="text/html" />
  <meta name="DC.type" scheme="DCTERMS.DCMIType" content="Text" />
</head>
```

Esimerkissä on käytetty tarkentimia. Esimerkiksi identifioinnin ominaisuuden tarkentimeksi on merkitty, että tunniste on URI-muodossa.

Seuraavaksi XML-kielinen esimerkki Dublin Core -määrittämisestä [10]. Määrittäminen ei sisällä tarkentimia.

```
<metadata
  xmlns="http://example.org/myapp/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://example.org/myapp/
    http://example.org/myapp/schema.xsd"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:title>UKOLN</dc:title>
  <dc:description>
    UKOLN is a national focus of expertise in digital information
    management. It provides policy, research and awareness services
    to the UK library, information and cultural heritage communities.
    UKOLN is based at the University of Bath.
  </dc:description>
  <dc:publisher>UKOLN, University of Bath</dc:publisher>
  <dc:identifier>http://www.ukoln.ac.uk/</dc:identifier>
</metadata>
```

Määrittäykset ovat osana laajempaa XML-sivua. dc-nimiavaruus erottaa metatietoa sisältävät elementit muista.

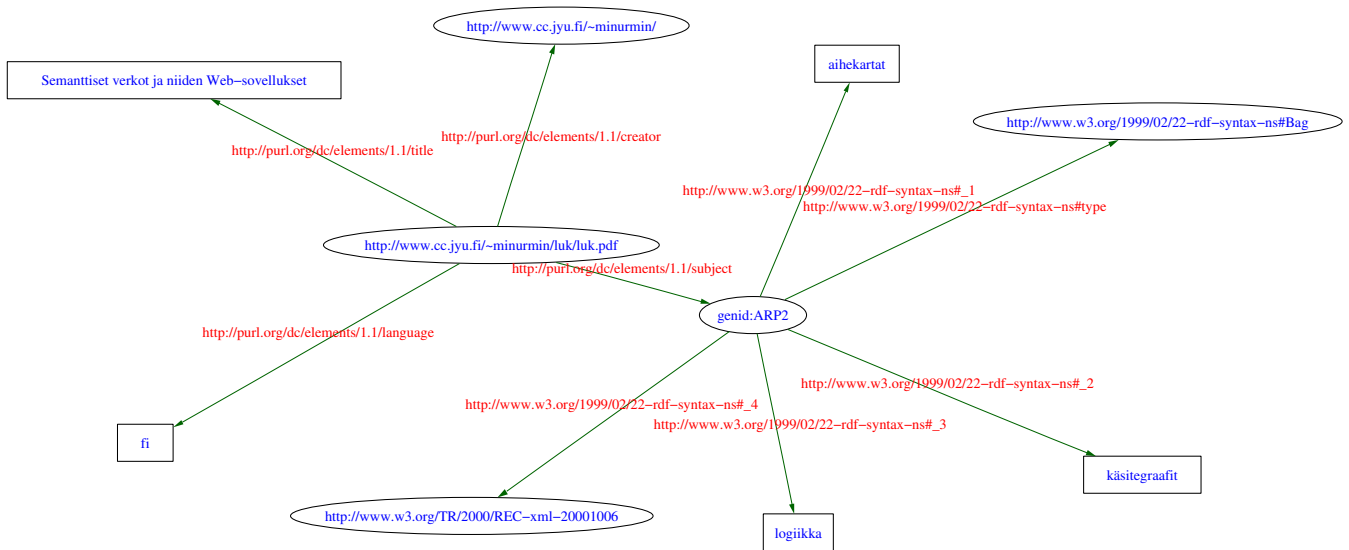
Viimeisenä esimerkkinä käsitellään RDF-kielinen esitys WWW-dokumentin ulkoisesta metatietokuvauksesta ilman tarkentimia. [8]

```

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description
    rdf:about="http://www.cc.jyu.fi/~minurmin/luk/luk.pdf">
    <dc:title>Semanttiset verkot ja niiden Web-sovellukset</dc:title>
    <dc:creator rdf:resource="http://www.cc.jyu.fi/~minurmin/" />
    <dc:subject>
      <rdf:Bag>
        <rdf:li>aihekartat</rdf:li>
        <rdf:li>käsittegraafit</rdf:li>
        <rdf:li>logiikka</rdf:li>
        <rdf:li
          rdf:resource="http://www.w3.org/TR/2000/REC-xml-20001006" />
          <!-- W3C:n XML -suositus -->
        </rdf:li>
      </rdf:Bag>
    </dc:subject>
    <dc:language>fi</dc:language>
  </rdf:Description>
</rdf:RDF>

```

Kuvauksessa käytetään edellisen esimerkin tapaan nimiavaruuksia siten, että RDF-kielen elementit ovat `rdf`-nimiavaruudessa ja Dublin Core -elementit ovat `dc`-nimiavaruudessa. RDF:n *bag*-tietorakennetta on hyödynnetty `subject`-elementit avainsanalistassa. Kuvassa 3 on esitetty RDF-kuvausta vastaava graafi. Kuva on luotu käyttäen W3C:n RDF-validaattoria⁶.



Kuva 3: RDF-graafi.

⁶<http://www.w3.org/RDF/validator/>

Lähteet

- [1] Asiakirjojen kuvailuformaatti. Tekninen raportti JHS 143, Julkisen hallinnon tietohallinnon neuvottelukunta, 2001. URL: <http://www.intermin.fi/intermin/hankkeet/juhta/home.nsf/pages/FCCC31E877F89F18C2256BED00291421>.
- [2] Dublin core metadata element set, version 1.1: Reference description. Tekninen raportti, DMCI Recommendation, 2003. URL: <http://dublincore.org/documents/dces/>.
- [3] T. Bray, D. Hollander ja A. Layman. Namespaces in xml. Tekninen raportti, W3C Recommendation, 1999. URL: <http://www.w3.org/TR/REC-xml-names>.
- [4] D. Hillmann. Using dublin core. Tekninen raportti, DMCI, 2003. URL: <http://dublincore.org/documents/usageguide/>.
- [5] E. Hyvönen, P. Harjula ja K. Viljanen. Representing metadata about web resources. Kirjassa E. Hyvönen, toim., *Semantic Web kick-off in Finland*, ss. 47–76. HIIT Publications, 2002. URL: <http://www.cs.helsinki.fi/u/eahyvone/stes/semanticweb/kick-off/proceedings.pdf>.
- [6] S. Kokkelin ja R. Schwänzl. Expressing qualified dublin core in rdf / xml. Tekninen raportti, DMCI Proposed Recommendation, 2002. URL: <http://dublincore.org/documents/dcq-rdf-xml/>.
- [7] F. Manola ja E. Miller. Rdf primer. Tekninen raportti, W3C Recommendation, 2004. URL: <http://www.w3.org/TR/REC-rdf-syntax/>.
- [8] M. Nurminen. Semanttiset verkot ja niiden web-sovellukset. LuK-tutkielma, Jyväskylän yliopisto, tietotekniikan laitos, 2003. URL: <http://www.mit.jyu.fi/luk/toteutettuja/SemanttisetVerkot/LuK.pdf>.
- [9] A. Powell. Expressing dublin core in html/xhtml meta and link elements. Tekninen raportti, DMCI Recommendation, 2003. URL: <http://dublincore.org/documents/dcq-html/>.
- [10] A. Powell ja P. Johnston. Guidelines for implementing dublin core in xml. Tekninen raportti, DMCI Recommendation, 2003. URL: <http://dublincore.org/documents/dc-xml-guidelines/>.
- [11] D. Raggett, A. L. Hors ja I. Jacobs. Html 4.01 specification. Tekninen raportti, W3C Recommendation, 1999. URL: <http://www.w3.org/TR/html401/>.
- [12] J. Stenvall ja J. Hakala. Dublin core -formaatin käyttöopas. Tekninen raportti, Helsingin yliopiston kirjasto, 1998. URL: <http://www.lib.helsinki.fi/meta/dc-opas.html>.
- [13] A. J. G. Swetland. Introduction to metadata: Setting the stage, 2000. URL: <http://www.getty.edu/research/institute/standards/intrometadata/>.
- [14] M. Vuori. Metatieto hyvinvoinnin ja terveyden indikaattorien kuvaamisessa. Tekninen raportti, STAKES, 2002. URL: <http://www.stakes.info/files/pdf/Hankkeet/Indikaattoripankki/metatietohvindi.doc>.