

## Interneto metaduomenys ir bibliotekinis katalogavimas

Juha HAKALA

Helsinki universiteto biblioteka - Suomijos nacionalinė biblioteka,  
P.O. Box 26, FIN-00014 Helsinki University, Finland, el. paštas: juha.hakala@helsinki.fi

### 1. Bendras metaduomenų apibūdinimas

Kiekvienas, kas nors kartą bandė surasti informaciją Internete, patyrė ne tik sėkmę, bet ir nusivylimą. Šių jausmų priežastį išaiškino Cliffordas Lynchas savo straipsnyje „Paieška Internete“ (<http://www.sciam.com/0397issue/0397lynch.html>):

Internetas, o ypač jo multimedia resursų fondas, žinomas kaip WWW, nebuvo sukurtas tam, kad padėtų organizuoti leidymus ar informacijos paieškai taip kaip bibliotekos. Jis išsivystė į tai, ką būtų galima pavadinti chaotišku kolektyvinės skaitmeninių „spausdinimo mašinų“ produkcijos sandėliu. Šis informacijos sandėlis apima ne tik knygas ir pranešimus, bet ir neapdorotus mokslu duomenis, meniu, protokolus, reklamą, garso ir vaizdo įrašus, interaktyvių pokalbių įrašus. Vienadienės reikšmės informacija susimaišiusi su išliekamąją vertę turinčiais darbais.

Tai leidžia C. Lynchui daryti išvadą:

Trumpai sakant, tinklas - tai ne skaitmeninė biblioteka. Bet jeigu jis ir toliau augs ir klestės kaip nauja komunikacijos priemonė, turės atsirasti kas nors labai panašus į tradicinės bibliotekos paslaugas, kad būtų galima organizuoti, padaryti prieinamą ir saugoti tinklo informaciją.

Viena iš tradicinių bibliotekinių paslaugų - išteklių aprašymas. Internete, kai reikia vengti vartoti nelabai paplitusius terminus šiame kontekste šnekama ne apie katalogavimą ir bibliografinius duomenis, o apie metaduomenis ir metaduomenų gamybą. Tačiau neleiskite, kad terminai suklaidintų: su metaduomenimis bibliotekos dirba jau labai seniai. Bibliotekos OPAC yra geras aukštos kokybės metaduomenų pavyzdys.

Metaduomenys dažnai apibūdinami kaip duomenys apie duomenis. Interneto kontekste metaduomenys yra struktūrizuotas požymių aprašymas objekto, panašus į dokumentą (DPO). Metaduomenų struktūra apibūdinama formatais, kurie vėlgi yra kilę iš katalogavimo taisyklių. Metaduomenys gali padėti vykdyti daugelį funkcijų - dokumento patalpimą, suradimą, dokumentavimą, vertinimą ir atranką. Šiuos veiksmus galutinai gali atlikti

vartotojai - žmonės arba jų kompiuteriai.

Deja, Internete metaduomenų nėra daug. Be to, jie sunkiai prieinami, yra prastos kokybės ir dažnai klaidinantys. Labai lengva „apgauti“ programas renkant duomenis iš Interneto, kai savi dokumentai aprašomi terminais, neturinčiais nieko bendra su turiniu.

Šiuo metu iškilo didelis poreikis pagerinti Internete esančių metaduomenų kokybę ir kiekį. Tačiau tai neįmanoma, jeigu nėra jų formato, kurį galėtų naudoti visi. Tokiu formatu sparčiai tampa Dublin Core metaduomenų elementų rinkinys. Yra nemažai kitų projektų, besiremiančių Dublin Core, kurie tobulina DC priemones ir aprašymą. Dėl to, kai kurie joje nedalyvaujantys bibliotekininkai jaučiasi nesaugūs. Taip pat jiems kelia nerimą neaiškūs santykiai tarp MARC ir Dublin Core, trūksta aiškaus supratimo, ką gi su pastaruoju daryti. Šiuo straipsniu, kurį parašė Dublin Core pradininkas, bandoma atsakyti į šiuos aktualius klausimus.

### 2. Metaduomenų formatai

Apie metaduomenis apskritai ir konkrečiai apie metaduomenų formatus yra daug parašyta. IFLA surinko gerą metaduomenų resursų, esančių tinkle, bibliografiją (žr. <http://www.nlc-bnc.ca/ifla/II/metadata.htm>). Šioje greitai besivystančioje srityje tinklo resursai yra geresni už spausdintus - beveik viskas, kas yra spausdinama popieriuje, pasensta greičiau nei pasiekia skaitytojus.

Tinkle tarp naudingų dokumentų yra metaduomenų formatų apžvalga, kurią išleido Europos Sąjungos projektas DESIRE. Šiame dokumente, kuris prieinamas adresu - <http://www.ukoln.ac.uk/metadata/desire/overview>, metaduomenų formatai skirstomi į tris grupes: paprasti, struktūrizuoti ir turtingi.

Paprasti formatai yra privatūs ir remiasi viso teksto indeksavimu. Geras pavyzdys yra Alta Vista formatas. Jį labai lengva naudoti, bet jis nelabai tinka informacijos paieškai. Nėra būdo atskirti, ar jūs ieškote Alberto Einšteino kaip dalyko, ar kaip autoriaus, arba Vilniaus - kaip leidimo vietos ar kaip dalyko. Privatus formato pobūdis reikiškiai

trejopai: formatas niekada netaps standartu, sunku surasti dokumentų apie jį ir neįmanoma daryti įtakos jo turiniui.

Turtingi formatai remiasi tarptautiniais standartais. Jie paprastai taikomi siaurose srityse, pavyzdžiui, bibliotekos naudojami MARC, kartografiškai - FGDC ir t.t. Norint išmokti ir naudoti turtingą formatą reikia didžiulių pastangų, tokius formatus labai sunku parduoti „kitoms sritims“. Pavyzdžiui, Suomijoje archyvai ir muziejai atsisakė FINMARC, nes šis netenkino jų poreikių, tuo tarpu jis akivaizdžiai tenkina daugumą, o gal ir visus bibliotekų poreikius. Laimė, Suomijos muziejams, bibliotekoms ir archyvams pavyko susitarti dėl pagrindinio elementų skaičiaus, kurį privalėtų turėti visos sistemos. Tai įgalina atlikti lygiagrečiai paiešką šių organizacijų bibliografinių duomenų bazėse, jeigu randamas bendras pasikeitimo formatas ir užklausk pateikimo būdus visoms sistemoms.

Struktūrizuoti formatai yra įdomus paprastų ir turtingų formatų kompromisas. Šie formatai yra santykinai nauji ir daugelis jų buvo sukurti naudojimui Internete. Kartais jie yra siaurai specifiški, tačiau taikymo atžvilgiu jie visiškai bendri. Pradžioje struktūrizuoti formatai nebuvo reikiamai standartizuoti, bet dabar keli jų - ir būtent Dublin Core metaduomenų elementų rinkinys - imami naudoti kaip Interneto standartai. Šie formatai leidžia atlikti paiešką pagal laukus, t.y. įmanoma sukurti vartotojo interfaissu „plyšiais“, pavyzdžiui, autoriui, antraštei ir dalykams. Jų struktūra paprastai nėra tokia sudėtinga kaip turtingų formatų.

Turtingi formatai dažniausiai skiriami informacijos tarpininkams, profesionalams, mokantiems tinkamai naudotis šia priemone. Struktūrizuotą formatą galėtų naudoti profesionalai, pavyzdžiui, bibliotekininkai, tačiau pagrindinis šių formatų sukūrimo motyvas yra sudaryti sąlygas autoriams ir leidėjams patiems aprašyti savo publikuojamus dokumentus Internete.

Nėra jokių būdų bibliotekininkui aprašyti net pačius relevančiausius Interneto resursus, nes šių dokumentų skaičius yra pernelyg didelis. Būtina atminti, kad bibliotekos turi lygiagrečiai tvarkyti tradicinius ir elektroninius dokumentus, neturėdamos šiam darbui jokių papildomų išteklių. Neatidėliotinai reikia pagalbos, ir šiuo atveju autoriai bei leidėjai privalo turėti šiam darbui reikiamą kvalifikaciją, kad būtų galima lengviau surasti tinklo dokumentus. Kaip pagalbinę priemonę bibliotekos gali naudoti šių resursų aprašus kaip katalogavimo pagrindą MARC formatu, žinoma, jeigu įmanoma konversija į MARC.

Suprantama, kad neturėdami reikiamų priemonių ir nurodymų autoriai nesugebės pateikti metaduomenų bibliotekoms patogiausią formą. Bibliotekos turi teikti atitinkamą paramą „mėgėjams“, dirbantiems su tokiais duomenimis. Kol duomenys yra „hobis“, tai nėra itin svarbu,

bet kai metaduomenys įteisinami, dalykai tampa rimtesni. Danijoje yra nacionalinis standartas, reikalaujantis, kad Dublin Core tipo metaduomenys būtų įtraukti į kiekvieną viešąjį valstybėje leidžiamą dokumentą. Kaip užtikrinti, kad metaduomenys, kuriuos pateikia tūkstančiai valstybės tarnautojų, būtų teisingi ir naudingi? To neįmanoma tinkamai padaryti, neįtraukus į šį darbą bibliotekų.

Turtingi ir struktūrizuoti formatai turi daug bendrų elementų. Be specifinių srities dalykų, beveik visada yra bendras pagrindas, tam tikras elementų rinkinys. Tai atveria dvi puikias galimybes. Pirma, turėtų būti įmanoma sukurti formatą, kuriame būtų tik šis pagrindinis aprašymo elementų rinkinys ir nieko daugiau. Antra, naudojant šį bendrą rinkinį ar kitas priemones, turi būti įmanoma visada konvertuoti metaduomenis iš vieno formato į kitą. Vienas iš didžiausių uždavinių kuriant skaitmenines bibliotekas yra sudaryti galimybes vartotojams vienu metu atlikti duomenų paiešką daugelyje įvairių organizacijų kuriamų duomenų bazių. Šis klausimas, nors ir labai įdomus, nenagrinėjamas šiame straipsnyje; čia sutelkiamas dėmesys į pirmąją temą.

### 3. Dublin Core metaduomenų elementų rinkinys

Pagal Dublin Core namų puslapį (žr. [http://purl.oclc.org/metadata/dublin\\_core/](http://purl.oclc.org/metadata/dublin_core/)) tai yra:

15-os metaduomenų elementų rinkinys, kurio paskirtis yra palengvinti elektroninių išteklių suradimą. Sukurtas pirmiausia autoriams, kūrėjams, kad jie patys galėtų aprašyti išteklius tinkle, jis taip pat patraukė tokių oficialiųjų išteklių aprašančių įstaigų kaip muziejai ir bibliotekos dėmesį.

Dublin Core sukūrė neformali bibliotekininkų grupė - tinklo ir turinio specialistai. Nuo to laiko, kai Dublin Core buvo sukurtas, t.y. nuo 1995 m. kovo mėnesį vykusio pirmojo metaduomenų seminaro, jis greitai tapo populiarus. Diegėjų grupė vis didėjo ir 1998 m. vasarą Dublin Core tapo Interneto standartu. Tolimesnis formato palaikymas bus taip pat formalizuotas, kad vartotojai galėtų žinoti, jog Dublin Core yra nuolat tobulinamas. Pagrindinės tobulinimo vietos ir toliau bus Dublin Core metaduomenų seminarai. Penktasis seminaras vyko Helsinkyje (žr. <http://linna.helsinki.fi/meta/DC5.html>). Šeštasis metaduomenų seminaras buvo surengtas Vašingtone, 1998 m. lapkričio mėnesį.

#### Elementai

Dublin Core elementų apibrėžimą galima gauti šiuo adresu - [http://purl.oclc.org/metadata/dublin\\_core\\_elements](http://purl.oclc.org/metadata/dublin_core_elements). Specifikacija išversta į daugelį kitų kalbų.

Dublin Core elementus galima padalinti į 3 klases (turinio, intelektualinės nuosavybės ir instancijos) šiuo būdu:

Turinys	Intelektualioji nuosavybė	Instancija
Antraštė	Kūrėjas	Data
Dalykas	Leidėjas	Tipas
Aprašas	Platintojas	Formatas
Šaltinis	Teisės	Identifikatorius
Kalba		
Ryšys		
Apimtis		

Daugeliu atvejų apibrėžti elementus yra santykinai paprasta. Pavyzdžiui, kūrėjas apibrėžiamas taip:

Asmuo ar organizacija, atsakinga už šaltinio intelektualiojo turinio sukūrimą. Pavyzdžiui, rašytinių dokumentų autoriai, vaizdinių dokumentų dailininkai, fotografai ar iliustratoriai.

Yra keli elementai, kurie nėra tokie akivaizdūs kaip kūrėjas. Vienas iš problemiškesnių elementų buvo Šaltinis, kuris apibrėžiamas taip:

Informacija apie antrinį šaltinį iš kurio paimtas esamas dokumentas. Nors apskritai rekomenduojama, kad elementuose būtų tik informacija apie esamą dokumentą, šiame elemente gali būti data, kūrėjas, formatas, identifikatorius arba kiti metaduomenys ir apie antrinį, nes manoma, jog tai svarbu esamo dokumento paieškai; geriausia būtų vietoje jo naudoti Santykio elementą. Pavyzdžiui, galima naudoti Šaltinio datą 1603, aprašant 1996 m. sukurtą filmą pagal Šekspyro pjesę, tačiau geriau būtų naudoti Ryšį „Remiasi“ su nuoroda į tam tikrą šaltinį, kurio apraše yra data 1603. Šaltinis nenaudojamas, jeigu esamas dokumentas yra pateikiamas originalia forma.

Buvo padarytas tvirtas sprendimas neįtraukti jokių naujų elementų ir nenaikinti senų. Jeigu kas nors nori įtraukti naują elementą, tai turi padaryti apibrėžtu būdu. Bet kuris privatus elementas turi prasidėti „X“. Pavyzdžiui, jei mes norime papildyti DC įrašą informacija apie kainą, tai reikia pridėti šešiolikto elementą „X - Kaina“. Tas pats mechanizmas gali būti naudojamas įvedant vietinius klasifikatorius (žr. toliau).

Kongreso biblioteka nuo pat pradžių buvo aktyvi Dublin Core seminarų dalyvė. Todėl nenuostabu, kad šiame lygyje Dublin Core ir MARC ėjo greta ta prasme, kad bet ką, įdėtą į „neklasifikuotą“ Dublin Core galima konvertuoti į MARC.

## Elementų rodikliai

Dublin Core su 15 elementų yra labai paprastas formatas. Vienas iš Dublin Core kūrėjų tikslų - padaryti formatą lankstų ir išplečiamą. Tai daroma naudojant elementų rodiklius.

Yra trys rodiklių rūšys: kalba, schema ir tipas (poelemento pavadinimas).

## Kalba

Kalba apibūdina elemento kalbos vertę. Jeigu pats dokumentas yra parašytas anglų kalba, tai gali būti nurodyta kalbos elemente. Tačiau jeigu dalykinės rubrikos yra pateikiamos anglų, estų ar lietuvių kalba, tai gali būti apibūdinama naudojant reikiama kalbos klasifikatorių kiekvienoje Dalyko žymoje. Apjungtas su naudingais bendro naudojimo bruožais HTML 4.0, pavyzdžiui, <HTML lang =“en“>, kalbos rodiklis žymiai išplės tinklo indeksų galimybes lokalizuoti, gauti ir indeksuoti dokumentus įvairiomis kalbomis.

## Klasifikavimo schema

Schema buvo apibrėžta kaip „formalus duomenų turinio standartas arba kodavimo standartas, turintis patikimas palaikymo priemonės“. Plačiai paplitę yra McSH (Medicinos dalykinės rubrikos), UDK (Universali dešimtainė klasifikacija) ir ISO 8601 standartas informacijos kodavimui. Schemos rodiklį apibūdina konkretaus elemento interpretavimo kontekstas. Tipiniu atveju tai yra nuoroda į tam tikrą schemą ar standartą. Pavyzdžiui, Dalyko laukas gali būti klasifikuotas SCHEME, kad jame būtų galima įrašyti UDK duomenis.

Yra atveju, kai schemos klasifikatorius turi didelę reikšmę lauko naudojimui. Pavyzdžiui, schema yra reikalinga datos interpretavimui. Datos nagrinėjimui reikia žinių apie kodavimo standartą (pavyzdžiui, schema), naudojamą išreiškiant turinį. Pavyzdžiui, ar eilutė 1998-05-11 reiškia lapkričio 5 d. ar gegužės 11 d.? Dėl tos pačios priežasties klasifikacija, įdėta į Dalyko žymę, turi būti apibūdinta naudojant schemą, nes priešingu atveju eilutė 681.03.06 (UDK indeksas programavimo kalboms) neturės prasmės taikymui, bandant jį analizuoti.

Data nekelia problemų, jeigu yra priemonių, kurios palaiko duomenų įvedimą, generuojant reikalaujamą HTML sintaksę. Dalykas yra daug sudėtingesnis, nes kontroliuojami tokie žodynai, kurie naudojami turinio aprašymui, tačiau tinkle jie paprastai nenaudojami. Negalima manyti, kad autoriai ir leidėjai įsigijo McSH ar UDK; šios sistemos turėtų būti prieinamos interaktyviai ir sujungtos su priemonėmis, naudojamomis kuriant Dublin Core įrašus. Informacijos paieškos požiūriu labai svarbu, kad metaduomenyse būtų

pateikiami ne tik bibliografiniai duomenys, bet ir aukštos kokybės turinio aprašymas, parengtas naudojant kontroliuojamus terminus. Dėl to metaduomenų kūrėjas, įtrauktas į Nordic Metadata projektą, maksimaliai palaiko dalykinę schemą, užtikrindamas tiesioginį ryšį su visomis žinomomis tinkle prieinamomis klasifikacijomis ir dalykinių rubrikų sąrašais (žr. kūrėjo Dalyko pagalbos puslapį <http://www.ub.lu.se/metadata/subject-hclp.html>).

## Tipas

Tipas rodiklis apibūdina duoto lauko facetą. Tipas rodiklio tikslas - susiaurinti elemento semantiką. Sudėtinga buvo parengti visuotinai priimtina tipo rodiklių sąrašą. Iki šiol nėra galutinio Dublin Core poelementų sąrašo, bet beveik pilną sąrašą galima gauti šiuo adresu - <http://lcweb.loc.gov/marc/dcqualif.html>.

Pagal šį sąrašą kūrėjas turi tokius „poelementus“:

- DC. Kūrėjas (be rodiklio)
- DC. Kūrėjas. Asmenvardis
- DC. Kūrėjas. Kolektyvas (įskaitant konferencijos pavadinimą)
- DC. Kūrėjas. Asmenvardis. Adresas (įskaitant visų tipų adresus, elektroninį pašta ir kitus)
- DC. Kūrėjas. Kolektyvas. Adresas

Įdomesnis atvejis yra Data, nes elektroninių dokumentų atveju data yra sudėtingas dalykas. Datai yra siūlomi (praktiškai jau patvirtinti) tokie rodikliai:

1. DC. Data. Turinio sukūrimas  
Intelektualaus turinio sukūrimo data
2. DC. Data. Variantų sukūrimas  
Resurso esama forma sukūrimo arba modifikavimo data
3. DC. Data. Skelbimas  
Formalaus dokumento skelbimo (pavyzdžiui, autorių teisių) data
4. DC. Data. Prieinamas  
Data, kai dokumentas taps arba tapo prieinamu
5. DC. Data. Galioja  
Dokumento galiojimo data (dažnai ribos)

## 6. DC. Data. Priimtas

Dokumento priėmimo (pavyzdžiui, disertacijos ar sutarties) data

## 7. DC. Data. Duomenys surinkti

Informacijos, esančios dokumente, surinkimo data

## 8. DC. Data. Įsigijimas

Įsigijimo ar katalogavimo data

Prašau atkreipti dėmesį, kad datos informacija kaip dalykas yra dedama į Dalyką ar Apimtį.

Iš pateikto sąrašo galime daryti išvadą, kad MARC formatu su datomis susijusi informacija nėra tinkamai tvarkoma. (Kitas MARC trūkumas yra tai, kad MARC turės metų skaičiavimo problemą 2000 metais.)

Šiuo metu Dublin Core turi tik 15 elementų, tačiau jį galima plėsti už MARC ribų, ir įgudęs vartotojas gali jį naudodamas parengti gerus dokumentų aprašus. Visa informacija, užkoduota naudojant MARC, gali būti konvertuojama į kokybišką Dublin Core įrašą, nors dėl savitos MARC įrašo struktūros konversija yra sudėtinga. Konversija iš MARC į DC buvo išbandyta vykdant Nordic Metadata projektą.

Projektai atliekami skirtingose sferose, pavyzdžiui, muziejai, bibliotekos, archyvai naudoja Dublin Core skirtingai, ir dėl to įrašai turi nedaug ką bendra. Bet Dublin Core vis dėlto yra mažiausias bendras šių skirtingų sferų bei projektų vardiklis ir užtikrina jų tarpusavio suderinamumą tiek, kiek tik tai įmanoma. Todėl nenuostabu, kad viena iš siūlomų Dublin Core naudojimo galimybių yra naudoti jį kaip pasikeitimo formatą. Taip naudojamas Dublin Core sudarys galimybes perduoti bibliografinius duomenis įvairioms sritims taip pat, kaip ISO 2709 (Tarptautinis pasikeitimo formatas) sudaro galimybes keistis MARC įrašais įvairių šalių bibliotekoms.

## Sintaksės klausimai

Metaduomenys gali būti vidiniai, t.y. aprašomo dokumento dalis, arba išoriniai. Metaduomenys, kuriami tarpininkų, yra išoriniai; vienas iš pavyzdžių - MARC duomenys (nors MARC duomenys taip pat gali būti dokumento dalis).

Dublin Core informacija paprastai yra įterpta pačiame dokumente. Tačiau tai nėra būtina; muziejai siekia naudoti išorinius Dublin Core metaduomenis. Bet jeigu Dublin Core yra įterpti į HTML, „Microsoft World“ ir kt., būtina sukurti Dublin Core sintaksę kiekvienam dokumento formatui (arba

įterpti aprašytą dokumentą kaip hiperryšį į HTML dokumentą, turintį Dublin Core dokumento aprašą).

Šiuo metu egzistuoja DC sintaksė HTML 2.0 ir HTML 4.0 (žr. [http://purl.oclc.org/metadata/dublin\\_core/syntax.html](http://purl.oclc.org/metadata/dublin_core/syntax.html)). Kitas žingsnis bus sintaksė XML Resursų aprašymo sistemai, RDF (žr. <http://www.w3.org/RDF/>). Nuo 1996 m. buvo diskutuojama dėl DC sintaksės sukūrimo plačiai naudojamiems vaizdo formatams, tokiems kaip PNG TIFF, tačiau iki šiol šioje srityje nieko nepadaryta. Iš esmės Dublin Core tinka vaizdo duomenų aprašymui.

Tarp Dublin Core kūrėjų vyrauja nuomonė, jog būtina įtraukti specifines priemones, kurios užtikrintų Dublin Core duomenis HTML standarte. Tai yra įmanoma dėl dviejų naujų HTML META žymės požymių, įvedus HTML 4.0 kaip Interneto standartą (žr. <http://www.w3.org/TR/REC-html40/>):

SCHEME: šis požymis įvardija schemą, kurią reikia naudoti interpretuojant savybės vertę, ir

LANG: kalbos žymėjimas remiantis RFC 1766 žymėmis kalbos identifikavimui.

Taigi galima sukurti, pavyzdžiui, tokią META žymę:

```
<META name="DC.Subject" Scheme = „YSA“
Lang="fi" Content= „luettelointi“>
```

arba tas pats anglų kalba:

```
<META name="DC.Subject" Scheme = „LCSH“
Lang="en" Content= „cataloguing“>
```

Pastaruoju metu tipo rodikliai HTML META žymėje buvo koduojami keliais būdais. HTML 2.0 efektyvi Dublin Core Autoriaus žymė atrodė taip:

```
<META NAME="DC.author" CONTENT=
„(TYPE=name) Hakala, Juha“>
```

Panašaus įrašo naudojant HTML 4.0 ir modernizuoto tipo rodiklio kodavimo priemones sintaksė yra tokia:

```
<META NAME="DC.Creator.PersonalName"
CONTENT="Koch, Traugott">
```

Pakankamai išsamus šio straipsnio Dublin Core įrašas HTML 4.0 atrodo taip:

```
<META NAME="DC.Title"CONTENT="Internet
metadata and library cataloguing">
```

```
<META NAME="DC.Creator.PersonalName"
CONTENT="Hakala, Juha">
```

```
<META NAME="DC.Creator.PersonalName.
Address"CONTENT="juha.hakala@helsinki.fi">
```

```
<META NAME="DC.Subject" CONTENT="Internet
resource description">
```

```
<META NAME="DC.Subject" CONTENT= „Dublin
Core metadata element set“>
```

```
<META NAME="DC.Description" CONTENT=
„This document describes the Dublin Core Metadata ele-
ment set and some Dublin Core projects.“>
```

```
<META NAME="DC.Publisher" CONTENT=
„Helsinki University Library“>
```

```
<META NAME="DC.Date" SCHEME= „ISO8601“
CONTENT="1998-05-24">
```

```
<META NAME="DC.Type" CONTENT="Text. Ar-
ticle">
```

```
<META NAME="DC.Format" SCHEME= „IMT“
CONTENT="text/html">
```

```
<META NAME="DC.Identifier" CONTENT= http://
linnae.helsinki.fi/meta/vilnius.html">
```

```
<META NAME="DC.Identifier" SCHEME= „URN“
CONTENT="URN:NBN:fi-fe19981038">
```

```
<META NAME="DC.Language" SCHEME=
„ISO639-1"CONTENT="en">
```

```
<META NAME="DC.Rights" SCHEME="Free text"
CONTENT="Public domain">
```

```
<META NAME="DC.Date.X-MetadataLastMo-
dified" SCHEME="ISO8601" CONTENT="1998-05-24">
```

Santrumpa „DC“ duomenų pradžioje rodo, kad buvo naudojamas Dublin Core. Tai būtina informacija automatiniam indeksavimui renkant iš dokumentų bibliografinius duomenis.

Būtina atsiminti, kad asmenų vardų norminimas įvedant duomenis (t.y. „vardas, pavardė“) yra ne bendro susitarimo rezultatas, o tik Nordic Metadata projekto rekomendacija. Ar vardai sunorminti, ar ne - nėra labai svarbu paieškos požiūriu (nors yra svarbu rūšiuojant rodykles), tačiau tai yra didelė MARC konversijos programų problema. Apskritai Dublin Core neturi nieko panašaus į tarptautinį bibliografinio aprašo standartą ir mažai tikėtina, kad turės ateityje. Rengiamas Dublin Core vartotojo vadovas šią problemą tam tikru mastu palengvins. Tačiau Dublin Core

vartotojų bendruomenė niekada nebus tokia glaudi kaip MARC vartotojų.

Iš pateikto įrašo matyti, kad Dublin Core yra paprastas formatas, tačiau jo HTML sintaksė gali būti gana sudėtinga. Be šio sudėtingumo programos, naudojančios Dublin Core, negalėtų suprasti duomenų semantikos. Pavyzdžiui, įrašą galima efektyviai indeksuoti ir konvertuoti į kitus formatus, įskaitant MARC. Kyla klausimas, kaip pasiekti, kad autoriai ir leidėjai gamintų tokius sudėtingus duomenis. Atsakymas: reikia naudoti tinkamas priemones. Nemokamai gaunamos Dublin Core priemonės, laimei, yra tobulinamos vykdamas daugelį projektų. Sutelkiame dėmesį į Nordic Metadata projektą.

### Nordic Metadata projektas

Projekto tikslas - sukurti priemones Dublin Core pagrįstiems metaduomenims kurti, rinkti ir indeksuoti. Projektą (žr. <http://linnae.helsinki.fi/meta>) iš dalies finansavo NORDINFO (<http://www.nordinfo.helsinki.fi/english/index.html>). Jis pradėtas vykdyti 1996 m. kaip pirmasis tarptautinis Dublin Core projektas, o baigtas 1998 m. gegužės mėnesį. Galutinę ataskaitą galima gauti adresu - <http://linnae.helsinki.fi/meta/nmfinal.htm>.

Projekte dalyvavo šios organizacijos ir asmenys:

Nordic Metadata dalyviai	
Bibsys, Norvegija	Ole Husby
Danijos bibliotekų centras, Danija	Susanne Thorborg
Helsinkio universiteto biblioteka, Suomija	Juha Hakala
Lundo universiteto biblioteka, NetLab, Švedija	Traugott Koch
Munksgaard, Danija	Anders Geertsen
Islandijos nacionalinė ir universiteto biblioteka, Islandija	Sigbergur Fridriksson
Švedijos kompiuterių mokslo institutas, Švedija	Preben Hansen

Projekto vadovas buvo Juha Hakala, parengęs ir pačią projekto idėją. Didelę įtaką projekto plano turiniui padarė Ole Husby ir Traugott Koch.

Metaduomenų priemonės ir adresai yra šie:

### Metaduomenų etalonas

([http://www.ub.lu.se/metadata/DC\\_creator.html](http://www.ub.lu.se/metadata/DC_creator.html))

Etaloną sukūrė Lundo universiteto bibliotekos NetLab skyrius naudodamas Nordic Metadata resursus. Etalonas paskelbtas 1997 m. sausio mėnesį, po to jis daug kartų buvo tobulinamas naudojant resursus, kuriuos NetLab gavo iš kitų metaduomenų projektų. Tobulinta remiantis Dublin Core pokyčiais ir vartotojų atsiliepimais.

Etalonas automatiškai kuria visą reikiamą HTML sintaksę ir leidžia vartotojui sutelkti dėmesį į turinio formavimą. „Perl“ šriftas etalonui kurti yra nemokamas. Jis buvo naudojamas keliuose kituose Šiaurės šalių projektuose (žr. galutinę ataskaitą). Mūsų statistika rodo, kad Dublin Core įrašus naudodami šį etaloną kūrė vartotojai iš daugiau kaip 30 valstybių. Vien Šiaurės šalyse yra daugiau nei 3000 Dublin Core įrašų. Tikimės, jog jų skaičius greitai didės, kai Dublin Core taps ne vien pomėgiu, bet ir profesionalumo dalimi.

Projektas taip pat parengė išplėtotus ir paprastus vartotojų vadovus Dublin Core metaduomenų gamintojams. Šie vadovai susiję su etalonu. Būtina ir toliau dirbti, nes norint, kad vadovai būtų tikrai naudingi vartotojams, juos reikia išversti ir modifikuoti, atsižvelgiant į vietos poreikius. Pavyzdžiui, Suomijos vadove turi būti informacija apie tai, kaip pateikti terminus iš Suomijos bendrojo dalykinių rubrikų sąrašo.

### Metaduomenų bazė

(<http://nwi.ub2.lu.se/?lang=en>)

Šiuo metu yra dvi metaduomenų bazės: Swemeta ir Danmeta, Švedijoje ir Danijoje. 1998 m. gegužės mėnesį jose buvo apie 250 000 įrašų, iš kurių tik nedidelė dalis buvo Dublin Core. 1998 m. vasarą buvo planuota sukurti metaduomenų bazes Suomijai ir Islandijai. Metaduomenų rinkimas ir indeksavimas vykdomas naudojant pagerintą Nordic Web Index programos versiją.

Metaduomenų bazės yra atskirtos nuo Nordic Web Index nacionalinių viso teksto duomenų bazių dėl aiškios priežasties - jeigu visi duomenys bus saugomi vicinteleje duomenų bazėje, bus prarandama pagrindinė metaduomenų teikiama nauda - didesnis paieškos tikslumas.

### Dublin Core konverteris į MARC

(<http://www.bibsys.no/meta/d2m/>)

Bibliotekų požiūriu yra svarbu turėti galimybę pritaikyti Dublin Core įrašus valstybinės bibliografijos duomenų bazių

ir bibliotekų OPAC palaikymui. Tam reikia Dublin Core konverterio į MARC.

Nordic Metadata projektas buvo pirmasis, sukūręs DC → MARC konverterį. Programą parengė Ole Husby iš Bibsys (Norvegija). Esama versija gerai funkcionuoja, bet palaikymo požiūriu ji turi esminį trūkumą: konversijos pokyčiai turi būti daromi modifikuojant programos kodą. Pradinis tikslas buvo padaryti lentelės tipo konverterį, bet darbo metu paaiškėjo, kad sukurti tokį konverterį yra daug sudėtingesnis uždavinys, negu manyta. Lentelės tipo konverteryje konversijos lentelė yra parengiama su, pavyzdžiui, Microsoft Excel, ir tada gauta elektroninė lentelė įvedama į konverterio programinę įrangą.

### URN generatorius

Santrumpa URN reiškia unifikuoti resursų vardai, leidžiantys identifikuoti Interneto dokumentus.

URN identifikatoriai yra pastovūs ir unikalūs: dokumentui suteiktas URN niekada nesikeičia, jeigu išlieka nepakitęs intelektualinis turinys. „Išnaudotas“ URN niekada nebus suteiktas kitam dokumentui.

URN sistemą sukūrė Internet Engineering Task Force's URN darbo grupė (žr. <http://www.ietf.org/html.charters/urn-charter.html>). 1998 m. gegužės mėnesį standartizavimas dar nebuvo baigtas, bet buvo sutarta dėl daugelio svarbių aspektų, pavyzdžiui, URN sintaksės.

Nacionalinių bibliotekų požiūriu vietinių Interneto dokumentų identifikavimas yra gyvybiškai svarbus, nes jis sudaro sąlygas efektyviai dokumentų paieškai ir saugojimui. Deja, esami bibliografiniai identifikatoriai gali būti naudojami tik nedidelei Interneto dokumentų daliai. Dėl šios priežasties vieta, arba URL, buvo dažnai naudojama identifikavimui, taip pat ir Dublin Core bendruomenėje, jeigu nebuvo galima gauti jokio tikro bibliografinio identifikatoriaus kaip ISBN ar ISSN.

URL niekada nebuvo kuriamas identifikavimo tikslais. Jis tik nurodo gaires, kaip pasiekti dokumentą (protokolas ir adresas, sudarytas iš įrengimo pavadinimo ir viso failo pavadinimo). Iš tikrųjų URL kaip identifikatorius teikia priešingus rezultatus: rinkimo programa vis tiek žino teisingą dokumento URL - priešingu atveju ji nesurastų dokumento, o jei dokumentas yra nukopijuotas kur nors kitur, senieji metaduomenys, jeigu jie netaisyti, yra klaidingi.

Nordic Metadata projektas siekė sudaryti vartotojams galimybę identifikuoti jų dokumentus kaip metaduomenų kūrimo proceso dalį. Todėl 1997 m. rudenį, kai URN sintaksė buvo užbaigta, pradėtos derybos su IETF URN darbo grupe dėl valstybinės bibliografijos numerių naudojamo URN. Rezultate sintaksė atrodo taip:

URN:NBN:xx-yy

Kiekvienas URN prasideda „URN:“; tai leidžia lokalizuoti URN iš paprastų dokumentų, neturinčių struktūros arba įterptų metaduomenų. NBN yra vardo vietos identifikatorius, arba NID valstybinės bibliografijos numeriams. Identifikuojanti URN dalis prasideda šalies kodu („xx“ aukščiau). Tai būtina, nes NBN („yy“ aukščiau) nėra vieningas visame pasaulyje - kiekviena nacionalinė biblioteka kaip NBN gali naudoti, ką ji nori.

Suomiškojo URN pavyzdys:

URN:NBN:fi-fe19981001

Lietuviškas URN turėtų prasidėti URN:NBN:lt-; likusi dalis priklauso nuo valstybinės bibliografijos naudojimo praktikos.

Kadangi NBN paprastai sudaro kodo informacija, metai ir numeris, NBN kūrimą lengva automatizuoti. Tai sakytina ir apie jais grindžiamas URN. Helsinkio universiteto biblioteka parašė specifikaciją URN generatoriui, kurį po to suprogramavo NetLab. Dabar programą galima gauti adresu - <http://www.lub.lu.se/metadata/URN-help.html>.

### Išvados

Bibliotekos jau seniai susidurdavo su metaduomenimis. Tačiau laikui bėgant metaduomenų gamybos priemonių nuolatos daugėja. OPAC ir MARC formatai neišnyksta, kaip ir spausdinti dokumentai neišnyks artimiausioje ateityje. Tačiau šalia tradicinių dokumentų biblioteka turi organizuoti bei daryti prieinamus taip pat ir elektroninius dokumentus, ir šiam tikslui kartu su tradiciniais metodais bus naudojamos ir naujos priemonės. Dublin Core metaduomenų elementų rinkinys ir yra viena iš naujųjų priemonių.

Šiuo metu galima įsigyti daug įvairių programų, kurias naudoja Dublin Core. Naudojant programinę įrangą, kuri buvo sukurta vykdant Nordic Metadata projektą, bibliotekos gali padėti vartotojams aprūpindamos juos metaduomenimis, o po to rinkti sukurtus metaduomenis iš tinklo ir sudaryti sąlygas vartotojams naudoti juos ieškant dokumentų. Vartotojai nebus suinteresuoti teikti metaduomenų, jeigu nepajus naudos.

Svarbu pažymėti, jog techninės priemonės nėra pakankamos metaduomenų palaikymui. Kadangi katalogavimo taisyklių nėra, vartotojams reikia metodinės medžiagos, kad būtų galima apibūdinti pageidaujamos rūšies Dublin Core. Taip pat būtina ir rengti vartotojus, ypač jeigu kataloguotojai nėra bibliotekininkai. Bibliotekininkai profesionalai, kaip rodo patirtis, lengvai perima Dublin Core, tačiau autoriams ir leidėjams atrodo, jog šis formatas - labai sudėtingas.

Galimybė stebėti Dublin Core atsiradimą per pastaruosius dvejus metus suteikė įdomios patirties. Nor-

dic Metadata ir kitos Skandinavijos metaduomenų tvarkymo iniciatyvos padėjo Šiaurės šalims gerai pasirengti aktyviai dalyvauti tolimesnėje veikloje. Jos yra tarp pirmųjų diegėjų šių naudingų naujų priemonių, kurias sukūrė Interneto metaduomenų bendruomenė.

Šiaurės šalys ir bibliotekos siekia ir toliau aktyviai rūpintis metaduomenų vystymu ir padėti savo kaimynams pradėti šį darbą. Visos priemonės, kurias sukūrė Nordic Metadata, yra ir bus nemokamai prieinamos, norint įdiegti jas Skandinavijoje ir kitose šalyse.