

RADIONICA DUBLIN CORE

Priredili:

Sofija Klarin, Laila Miletić-Vejzović (voditeljica), Damir Pavelić,
Sonja Pigac, Mirna Willer i Goran Zlodi

Izveštaj i zaključci

Cilj je radionice bio razmotriti relevantnost upotrebe metapodataka u elektroničkoj građi općenito, te upoznati upotrebu Dublin Corea na konkretnom primjeru. Pri tome se pod pojmom elektronička građa razumijevala izvorno nastala digitalna građa, kao i ona nastala postupkom digitalizacije.

U prvom je dijelu radionice Laila Miletić-Vejzović¹ govorila o upotrebi Dublin Corea kao formata za bilježenje bibliografskih podataka u opisu digitaliziranih fotografija. U tom je postupku najprije trebalo definirati upotrebu elemenata metapodataka Dublin Corea za pojedinu vrstu bibliografskog podatka. Npr. fotograf Frank S. Matsura označen je elementom DC.Stvaratelj, a Knjižnica države Washington elementom DC.Nakladnik. Definirane su i standardne vrijednosti pojedinih elemenata, npr. *copyright* kao u DC.Prava, format digitalizirane fotografije kao u DC.Format. Maska za unos podataka nije sadržavala izvorne nazive elemenata metapodataka Dublin Corea, već su oni dobivali karakteristične nazive za vrstu podatka koji se u tom projektu unosio u element Dublin Corea (*mapping*). Na primjer, spomenuti element DC.Stvaratelj preimenovan je u Fotograf, jer mu je i namjena da zabilježi upravo ime fotografa. (Taj će se element nazvati Kartograf u idućem projektu digitalizacije karata.) Nadalje, određeni su elementi Dublin Corea bili definirani kao obvezni odnosno neponovljivi unatoč tome što sam Dublin Core ne propisuje takva ograničenja. Posebno je detaljno L. Miletić-Vejzović opisala unos bibliografskih podataka. Premda je Dublin Core zamišljen kao format koji se može koristiti bez poznavanja pravila bibliotečne struke tj. deskriptivne i sadržajne katalogizacije, pri unosu su se podataka koristile prednosti oba ova pristupa. Iskorištena je mogućnost što Dublin Core ne propisuje određeni redoslijed i pravila pri unosu sadržaja metapodataka pa se broj elemenata metapodataka (ponovljivi elementi metapodataka) i njihov sadržaj mogao podvrgnuti osnovnoj svrsi ovog projekta: opisati fotografiju tako da korisnik dobije odgovor i na najdetaljniji upit. Npr. okrug – naziv mjesta – vrsta zgrade – naziv zgrade – vremensko razdoblje. Pravila Predmetnog sustava Kongresne knjižnice koja se inače koriste pri sadržajnoj obradi tradicionalne građe, ne dopuštaju takav redoslijed podataka, odnosno takvo fragmentiranje sadržajnog opisa koji omogućuje najrazličiće kombinacije pojmova pri pretraživanju. Međutim, pri upotrebi predmetnih pojmova koristio se gdje god je bilo moguće predmetni

¹ O projektu digitalizacije Zbirke fotografija Franka S. Matsure vidjeti u radu Laile Miletić-Vejzović u ovom Zborniku.

tezaurus i normativni oblici imena i naziva spomenutog pravilnika (npr. ime fotografa za-bilježeno je u obliku: Matsura, Frank S.).

L. Miletić-Vejzović prikazala je također rad s programom CONTENT Database i dala popis mrežnih adresa važnih za digitalizaciju i označivanje digitalnih dokumenata metapodacima (Prilog 1).

U drugom su dijelu radionice Sofija Klarin i Sonja Pigac prikazale rezultate istraživanja o daljinski dostupnim hrvatskim serijskim publikacijama i njihovoj kataložnoj obradi u su-stavu CROLIST Nacionalne i sveučilišne knjižnice. Nastavak tog istraživanja autorice su vidjele u ispitivanju metapodataka ugrađenih u elektroničke serijske publikacije (Prilog 2). U daljnji rad uključio se i Damir Pavelić koji je podrobnije govorio o metapodacima elek-troničkih serijskih publikacija općenito, njihovoj upotrebi te standardima XML i RDF. Zaključno, D. Pavelić istaknuo je promjenu u načinu korištenja metapodataka, od njihove prvotne upotrebe kao elemenata opisa elektroničkog dokumenta radi njegove pretražljivo-sti na internetu do njihove uloge pri razmjeni podataka. Pritom je od presudne važnosti kako za hrvatske knjižnice tako i za druge ustanove da aktivno prate razvoj standarda i započnu koristiti Dublin Core te RDF i XML u svojim projektima (Prilog 3).

U trećem se dijelu radionice, koji su vodili Mirna Willer i Goran Zlodi, raspravljalo o elementima metapodataka Dublin Corea na osnovi teksta *CIMI Guide to Best Practice: Du-blin Core* što ga je prevela Markita Franulić.² Zaključeno je da je načelno nužna upotreba standarda pri opisu građe (sadržaj – elementi podataka i njihove vrijednosti tj. pravila koja propisuju oblik elemenata podataka) i načinu njena bilježenja (označitelji sadržaja (seman-tika) – Dublin Core i “tvarni” nositelji formata (sintaksa) – XML i RDF), da je upotreba Dublin Corea preporučljiva za opis daljinski dostupne elektroničke građe kao i digitalizira-ne građe, posebno ako se taj format može koristiti u bazi podataka, te da bi bilo potrebno pristupiti izradi *mappinga* Dublin Corea za upotrebu u konkretnim projektima arhiva, knjižnica i muzeja, ponajprije radi transparentnog (*seamless*) pretraživanja njihovih baza podataka. Ovi su zaključci ujedno prihvaćeni i kao zaključci radionice.

² Franulić, Markita. Dublin Core u muzejskoj zajednici. // *Informatica museologica* 31, 1/2(2000), 16-24.

WORKSHOP DUBLIN CORE

Prepared by:

Sofija Klarin, Laila Miletić-Vejzović, Damir Pavelić, Sonja Pigac,
Mirna Willer and Goran Zlodi

Report and conclusions

The aim of this workshop was to study the uses of metadata in electronic material in general, and to become acquainted with the use of Dublin Core as a concrete example. Here the term electronic material includes originally created digital material, and material created through digitisation.

In the first part of the workshop Laila Miletić-Vejzović¹ spoke of the use of Dublin Core as the format for recording bibliographic data in the description of digitised photographs. In this procedure the use of Dublin Core metadata for a specific type of bibliographic element has to be defined. For example, a photographer Frank S. Matsura was defined with the element DC.Creator, and Washington State Library with the element DC.Publisher. The standard values of individual elements were defined as well, e.g. copyright as in DC.Rights, the format of digitised photographs as in DC.Format. The interface for data entering did not contain the original names for Dublin Core metadata elements, but were given characteristic names for the types of data which that project used for the Dublin Core element (mapping). For example, the element DC.Creator was renamed Photographer, because its purpose is to record the name of a photographer. (In the future project on map digitisation this element will be called Cartographer). Furthermore, certain Dublin Core elements were defined as obligatory, i.e. unrepeatable, even though Dublin Core prescribes no such limitation. Particularly detailed was L. Miletić-Vejzović's description of recording bibliographic data. Even though Dublin Core was designed as a format that can be used without librarianship knowledge, i.e. descriptive and subject cataloguing, in recording data both approaches were used. The fact that Dublin Core does not prescribe any specific sequence or rules for recording data, so the number of metadata (repeatable metadata elements) and their content were able to be submitted for the basic purpose of this project: to describe a photograph in such a way that a user can obtain an answer even to the most detailed query. For example, district-name of the place-type of the building-time period. The rules of the Library of Congress Subject Heading system, which are regularly used in subject cataloguing of traditional material, do not permit such sequence of data, i.e. such fragmenting of content description, which enables a wide variety of combinations of terms in searching. However, in the use of subject terms, wherever it was possible, subject thesaurus and authority forms for names and title using the aforementioned rules were used (e.g. the name of photographer was recorded in the form: Matsura, Frank S.).

¹ See Laila Miletić-Vejzović's paper on the project of digitization of the Frank S. Matsura's photograph collection in these proceedings.

L. Miletić-Vejzović also presented the work of the CONTENT Database program and gave a list of internet addresses relevant for digitisation and denoting digital documents with metadata (Appendix 1).

In the second part of the workshop Sofija Klarin and Sonja Pigac presented the results of their research on remotely accessible Croatian serial publications, and the results of the cataloguing project on the CROLIST system of the National and University Library. The authors described the continuation of this research in examining metadata embedded in electronic serial publications (Appendix 2). Following this D. Pavelić spoke in greater detail on electronic serial publication metadata generally, its use, and the XML and RDF standards. In conclusion, D. Pavelić stressed the developments in the use of metadata, from its original use in the description of electronic documents for retrievability on the internet, to its role in information exchange. In this process it is extremely important, equally for Croatian libraries as for other institutions, to actively follow the development of standards and to begin using Dublin Core, RDF and XML in their projects (Appendix 3).

In the third part of the workshop, led by Mirna Willer and Goran Zlodi, Dublin Core metadata elements were discussed on the basis of the text *CIMI Guide to Best Practice: Dublin Core* translated by Markita Franulić.² It was concluded in principle that the use of standards is necessary for material description (content-data elements and their value, i.e. the rules which prescribe the form of data elements) and for the way in which they are recorded (content designators (semantics) - Dublin Core, and "material" carriers of format (syntax) - XML and RDF). It was recommended that Dublin Core should be used for the description of remotely accessible electronic material and other digitised material (especially if that format can be used in a database), and that it would be necessary to define the mapping of Dublin Core for use in concrete projects in archives, libraries and museums, primarily for seamless searching of their databases. These recommendations were at the same time accepted as the conclusions of the workshop.

² Franulić, Markita. Dublin Core u muzejskoj zajednici. // *Informatica museologica* 31, 1/2(2000), 16-24.

PRILOG 1

**POPIS MREŽNIH ADRESA VAŽNIH ZA
DIGITALIZACIJU I OZNAČIVANJE DIGITALNIH
DOKUMENATA**

(1) WSU (Washington State University) Libraries CONTENT Server

WSU Libraries CONTENT Server

<http://content.wsulibs.wsu.edu/>

WSU Buildings Database - CONTENT HTML Client

<http://content.wsulibs.wsu.edu/cgi-bin/go.exe?CISOROOT=/buildings>

WSU Buildings Database - CONTENT Contextual Client

<http://www.wsulibs.wsu.edu/holland/masc/xbldingsfinal.html>

WSU Buildings Database - Collection Administration [Requires entry of authentication information]

<http://content.wsulibs.wsu.edu/cgi-bin/admin/home.exe?CISODB=/buildings>

Frank S. Matsura Database - CONTENT HTML Client

<http://content.wsulibs.wsu.edu/cgi-bin/go.exe?CISOROOT=/Matsura>

Frank S. Matsura Database - CONTENT Contextual Client

<http://www.wsulibs.wsu.edu/holland/masc/xmatsura.html>

Frank S. Matsura Database - Collection Administration [Requires entry of authentication information]

<http://content.wsulibs.wsu.edu/cgi-bin/admin/home.exe?CISODB=/Matsura>

Test Matsura Database - CONTENT HTML Client

<http://content.wsulibs.wsu.edu/cgi-bin/go.exe?CISOROOT=/mattest>

Test Matsura Database - Collection Administration [Requires entry of authentication information] Collection Administration [Requires entry of authentication information]

<http://content.wsulibs.wsu.edu/cgi-bin/admin/home.exe?CISODB=/mattest>

(2) *Referentni standardi*

[DCT1] DC Type Qualifiers. Dublin Core Draft Working Draft, 19 May 2000.
<http://lcweb.loc.gov/marc/dc/typequalif-20000519.html>

[ISO3166] ISO 3166 - Codes for the representation of names of countries.
<http://www.din.de/gremien/nas/nabd/iso3166ma/>

[ISO639] ISO 639-2 - Codes for the representation of names of languages, Alpha-3 code (ISO 639-2:1998). <http://www.loc.gov/standards/iso639-2/langhome.html>

[MIME] Internet Media Types.
<http://www.isi.edu/in-notes/iana/assignments/media-types/>

[RFC1766] Tags for the Identification of Languages, Internet RFC 1766.
<http://www.ietf.org/rfc/rfc1766.txt>

[RFC2396] Uniform Resource Identifiers (URI): Generic Syntax, Internet RFC 2396.
<http://www.ietf.org/rfc/rfc2396.txt>

[RFC2413] Dublin Core Metadata for Resource Discovery. Internet RFC 2413.
<http://www.ietf.org/rfc/rfc2413.txt>

[RFC2731] Encoding Dublin Core Metadata in HTML. Internet RFC 2731.
<http://www.ietf.org/rfc/rfc2731.txt>

[TGM1] Thesaurus for Graphic Materials II: Genre & Physical Characteristics Terms
<http://lcweb.loc.gov/rr/print/tgm2/>

[TGM1] Thesaurus for Graphic Materials I: Subject Terms
<http://lcweb.loc.gov/rr/print/tgm1/>

[TGN] Getty Thesaurus of Geographic Names.
http://shiva.pub.getty.edu/tgn_browser/

[W3C] The World Web Consortium
<http://www.w3.org/>

[W3CDTF] Date and Time Formats, W3C Note.
<http://www.w3.org/TR/NOTE-datetime-970915.html>

[XML] Extensible Markup Language.
<http://www.w3.org/XML/>

Pripremila: Laila Miletić-Vejzović

PRILOG 2

HRVATSKE SERIJSKE PUBLIKACIJE NA INTERNETU I METAPODACI

Sofija Klarin, Sonja Pigac

Nacionalna i sveučilišna knjižnica, Zagreb

1. Analiza hrvatskih daljinski dostupnih serijskih publikacija

Suočeni s izazovima tehnologije hrvatski nakladnici i tiskari posljednjih godina polako počinju primjenjivati nova tehnološka rješenja i uključivati se u elektroničko tiskarstvo i nakladništvo. Određeni se broj nakladnika, prateći trendove u svjetskom nakladništvu, prilagodio promjenama u načinu poslovanja izlaskom na internet što im je omogućilo proširenje tržišta i nove načine poslovanja. Osim tridesetak hrvatskih nakladnika koji se mrežom koriste za elektroničku promociju ili distribuciju svojih tiskanih publikacija,¹ javlja se manji broj onih koji su se započeli baviti elektroničkim nakladništvom, oblikom nakladništva u kojemu su publikacije proizvedene i uskladištene elektronički, a moguć im je i pristup putem računalnih mreža. Još uvijek je vrlo mali broj hrvatskih nakladnika koji su se okrenuli izdavanju isključivo elektroničkih publikacija, a ne samo digitaliziranih pretilisaka svojih tiskanih izdanja.² Zamjećujemo i pokušaje pojedinaca da u "vlastitoj nakladi" objavljuju na mreži,³ a posebno je zanimljiva pojava većeg broja serijskih publikacija. Tako na internetu nalazimo elektronička izdanja dnevnih novina (*Vjesnik*, *Večernji list*, *Slobodna Dalmacija*) i tjednika (*Nacional*, *Feral Tribune*, *Gloria*), nekoliko znanstvenih časopisa (*Acta pharmaceutica*, *Croatian Medical Journal Online*, *Paediatrica Croatica*) i niz isključivo elektroničkih naslova popularnih magazina, tzv. *e-zina* (*Internet Monitor*, *Klik*, *Netplugged*). Brojnost i promjene u načinu prezentacije, strukturiranja i izlaženja takvih daljinski dostupnih publikacija potaknule su nas da u razdoblju od 1. siječnja do 1. rujna 2000. godine istražimo i analiziramo hrvatske serijske publikacije na internetu.⁴ U tom smo istraživanju došle do sljedećih rezultata:

¹ Po oglašavanju i distribuciji knjiga putem Interneta najpoznatiji je nakladnik *Moderna vremena* - <http://www.moderna-vremena.hr>.

² Npr., nakladnik *Strijelac* iz Zagreba, <http://www.znak.hr>.

³ Npr., zbirka poezije *Riči* Berezine Matoković. Upravo na mreži izlazi i prva zbirka pripovijesti isključivo u elektroničkom obliku - *Iskonabule* hrvatskog književnika Roberta Perišića, <http://www.klik.hr/iskonabula>.

⁴ Opširniju analizu vidjeti u članku: Klarin, Sofija; Sonja Pigac. Hrvatske daljinski dostupne elektroničke serijske publikacije. // *Vjesnik bibliotekara Hrvatske* 43, 4(2000), 156-167.

Broj hrv. e-ser. publikac.	Prestale izlaziti	Nestale s mreže	Promijenile adresu	Promijenile naslov	Nisu javno dostupne	Jedinice s metapodacima
140	4	8	2	1	2	32,8 %

Vrsta domene	Izdanja	Tipologija	Kronologija izlaženja	Nakladnici e-ser. publ.	Jedinice s ISSN-om
.hr 81% .com 17% ostalo 1%	samo el. 35 el. i tisk. 105	novine 24 časopisi 108 ostalo 8	1994 - 1 1995 - 7 1996 - 11 1997 - 21 1998 - 26 1999 - 33 2000 - 20 nepoznato - 21	Zagreb - 109 Split - 6 Rijeka - 3 Osijek, Dubrovnik, Varaždin, Čakovec - 2 Karlovac, Zadar, Slavonski Brod, Pula, Koprivnica, Ičići, Prelog, Sv. Ivan Zelina - 1 drugo - 1 nepoznato - 4	14

Analiza je pokazala da u ukupnom broju od 140 hrvatskih daljinski dostupnih serijskih publikacija dominira velik broj časopisa (od čega 7 znanstvenih), da 25% jedinica izlazi samo na mreži. Podatak da je u 8 mjeseci nestalo 5,7% jedinica upozorava na potrebu što hitnijeg početka pohrane i zaštite takve građe u prikladnom nacionalnom depozitoriju.

2. Metapodaci

U okviru istraživanja pregledani su kodovi svih html dokumenata i načinjena je analiza zastupljenosti metapodataka. Rezultat od 32,8% publikacija s metapodacima mogao bi se učiniti velikim, ali kvaliteta metapodataka pokazuje da hrvatski nakladnici serijskih publikacija nisu dovoljno upoznati s razlozima uporabe i sustavima metapodataka za mrežne publikacije, a u određenim slučajevima upitan je i način njihove uporabe.

Primjer 1 Metapodaci e-zina *Internet Monitor*, <http://www.monitor.hr>

```
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=windows-1250">
<meta name="author" content="Internet Monitor">
<meta name="keywords" content="Internet Monitor Hrvatska supermodels software
muzika teatar cool kultura pop">
<meta name="description" content="Internet Monitor pop kulturni e-zine">
<title>Internet Monitor</title>
```

Primjer 2 Metapodaci časopisa *Banak*, <http://www.banak.net>

```
<head>
<meta http-equiv="Content-Type" content="text/html; charset=windows-1250">
<meta name="description" content="BANAK je najtiražnija revija za marketing u Dal-
maciji. Najveći broj komercijalnih i besplatnih malih oglasa na jednom mjestu. Svaki tje-
dan novi mali oglasi na http://www.dalmatianet.com/banak istovremeno s papirnatim iz-
danjem na kioscima.">
<meta name="keywords" content="oglas, mali oglasi, hrvatska, croatia, dalmacija, dal-
matia, split, marketing, tjednik, revija, besplatno, besplatni, slododna, slobodna dalmacija,
stan, stanovi, posao, business, zaposlenje">
<meta name="GENERATOR" content="Microsoft FrontPage 4.0">
<title>BANAK Dalmatinska revija za marketing - besplatni mali oglasi svaki tje-
dan!</title>
</head>
```

Primjer 3 Metapodaci časopisa *ACS*, <http://www.agr.hr/smotra>

```
<HEAD>
<META NAME="GENERATOR" CONTENT="Adobe PageMill 3.0 Win">
<META HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=iso-
8859-2">
<TITLE>ACS-AGRICULTURAE CONSPECTUS SCIENTIFICUS</TITLE>
LINK REV="made" HREF="mailto:smotra@agr.hr">
<META NAME="keywords" CONTENT="Croatia, agriculture, science, publication,
agricultural, economics, rural, sociology, plant, pathology, herbology, animal, nutrition,
engineering, soil, amelioration, microbiology, dairy, agronomy, breeding, genetics, botany,
zoology, crops, fishery, beekeeping, husbandry, forades, grassland, ornamental, landscape,
architecture, farm, management, enology, viticulture, pomology">
<META NAME="description" CONTENT="On-line Scientific Journal"
AGRICULTURAE CONSPECTUS SCIENTIFICUS PUBLISHED BY FACULTY OF
AGRICULTURE UNIVERSITY ZAGREB>
<META NAME="copyright" CONTENT="ACS Agriculture Conspectus Scientifi-
cus">
<META NAME="revisit-after" CONTENT="60 Days">
<META NAME="Robot" CONTENT="ALL">
<META NAME="DC.Title" CONTENT="ACS-AGRICULTURAE CONSPECTUS
SCIENTIFICUS">
<META NAME="DC.Creator" CONTENT="Agriculture Conspectus Scientificus, Fa-
culty of Agriculture, Zagreb CROATIA">
```

```
<META NAME="DC.Publisher" CONTENT="Faculty of Agriculture University of Zagreb">
```

```
</HEAD>
```

Primjeri pokazuju da su nakladnici u zaglavlju publikacija upotrijebili samo nekoliko osnovnih elemenata metapodataka, kao što su npr. podaci o naslovu, autoru, nakladniku, opis sadržaja. Čitav skup elemenata izrađen prema jednom od aktualnih sustava (Dublin Core, TEI, EAD i dr.) nije pronađen ni u jednoj publikaciji.

2.1. Zašto su metapodaci važni za elektroničko nakladništvo?

Obično se ističe da metapodaci imaju važnu ulogu u identifikaciji, pronalaženju, osiguravanju autentičnosti, pristupu i dugoročnoj zaštiti mrežne građe. Autori i nakladnici mogu doprinijeti povećanju pristupa svojim publikacijama tako da prije njihove objave na mreži izrade strukturirani opis, odnosno metapodatke publikacije rabeći jedan od sustava metapodataka. Metapodaci se mogu razmjenjivati u različite svrhe, ali je razmjena među samim nakladnicima i distributerima često komplicirana zbog različitih formata, odnosno nedostatka standarda. S obzirom da je i većim nakladnicima, kao i netransparentnoj i neformalnoj skupini pojedinaca autora-nakladnika, na World Wide Web-u teško nametnuti bilo kakve standarde, osim onih u kojima nalaze određen komercijalni interes,⁵ pokušava im se ukazati na važnost uporabe metapodataka za poboljšanje pristupa, promociju, prodaju pa i dugoročnu zaštitu njihovih publikacija. Ne treba posebno naglašavati da se metapodaci mogu efikasno koristiti, razmjenjivati i prenositi,⁶ kao uostalom i svi drugi strukturirani podaci, isključivo u standardiziranom formatu, npr. XML-u.⁷

Upravo se u tome nazire važnost uloge knjižničara u uspostavi suradnje s nakladnicima i autorima elektroničkih publikacija i pružanju pomoći, edukacije i alata za izradu strukturiranog opisa pomoću jednog od sustava metapodataka.⁸

2.2. Zašto su metapodaci važni za knjižnice?

Knjižničarima je važna identifikacija kvalitetnih publikacija na mreži kako bi svojim korisnicima mogli ponuditi i takvu vrstu izvora informacija. U identifikaciji im pomažu standardizirani metapodaci i pretraživači koji će ih moći detektirati na mreži. Osim toga, katalogizatori kvalitetne metapodatke mogu upotrijebiti pri obradbi internetske građe u postupku njihova automatiziranog preuzimanja i potom dorađivati rabeći standarde za meta-

⁵ Takav je uostalom bio slučaj i s identifikatorima publikacija kao što su ISBN, ISSN, ISMN, a tek ćemo vidjeti kakav će biti stav prema DOI-u, URN-u i drugim sustavima jedinstvenih identifikatora.

⁶ Vidjeti projekt INDECS (Interoperability of Data in E-Commerce Systems) čiji je cilj uspostava okvira za elektroničko poslovanje intelektualnim pravima u svim medijima, a osnovni produkt je model metapodataka za tu svrhu, <http://www.indecs.com>, kao i standard ONIX - Online Information eXchange, produkt udruga američkih i britanskih nakladnika za izradu metapodataka u XML-u, <http://www.editeur.org/onix.html>.

⁷ Vidjeti Prilog 3.

⁸ Vidjeti projekt BIBLINK, europski projekt suradnje depozitarnih knjižnica i nakladnika <http://hosted.ukoln.ac.uk/biblink>.

podatke (npr. Dublin Core) ili tradicionalne standarde svjetske knjižnične zajednice za opis elektroničke građe (npr. ISBD(ER), MARC/UNIMARC).⁹ Nacionalnim je knjižnicama i bibliografskim središtima važna identifikacija nacionalne produkcije elektroničke građe koju je, kao i sve druge vrste građe, važno obraditi, pružiti joj pristup, osigurati dugoročnu dostupnost. To se može osigurati na način da nakladnici registriraju svoje elektroničke publikacije u nacionalnoj knjižnici/bibliografskom središtu i pritom dostavljaju metapodatke. S obzirom na iskustva pri izradi bibliografskih zapisa, knjižničari mogu pomoći nakladnicima pri izradi i doradi kvalitetnih metapodataka.

Knjižnice posljednjih godina postaju i same nakladnici elektroničkih publikacija digitalizacijom dijelova svojih fondova u čemu su često zastupljena i godišta starijih serijskih publikacija koje valja zaštititi od uporabe, a korisnicima pružiti pristup elektroničkom obliku.¹⁰ U opisu digitaliziranih časopisa koji postaju tekstualne datoteke važna je uporaba metapodataka za opis cjeline i dijelova publikacije u širinu i dubinu. Granularnost časopisa sastavljenog od godišta, brojeva i članaka zahtijeva granularnost njegova zapisa koja se često u velikim projektima retrospektivne digitalizacije ne može osigurati izradom bibliografskih zapisa, bilo zbog nedostaka knjižničnog sustava koji ne podržava hijerarhijske odnose ili nedostatka sredstava i osoblja sposobnog za takve poslove. Kvalitetan i siguran pristup podacima može se, dakle, osigurati izradom metapodataka članaka.¹¹ Pritom se ne smije zaboraviti ni standard za identifikatore dijelova publikacija pa tako i elektroničkih članaka.¹² I na kraju, u knjižnicama će se sve više govoriti o tzv. tehničkim metapodacima koji se izrađuju u svrhu dugoročne zaštite građe i osiguravaju uspješnu migraciju dokumenata bez većih gubitaka.¹³

⁹ Kao na primjer u Projektu OCLC-a CORC - Cooperative Online Resource Catalog, <http://www.oclc.org/oclc/corc/> gdje se u kooperativnoj katalogizaciji daljinski dostupne elektroničke građe izrađuju tradicionalni kataložni zapisi kao i metapodaci u Dublin Core formatu.

¹⁰ Vidjeti projekte digitalizacije časopisa DIEPER - Digitised European periodicals, <http://www.sub.uni-goettingen.de/gdz/dieper/> i JSTOR - Journal Storage, <http://www.jstor.org/>.

¹¹ Izrada metapodataka na razini članka izuzetno je važna za distribuciju, odnosno dostavu ili tiskanje na zahtjev.

¹² SICI: Serial Item and Contribution Identifier Standard, <http://sunsite.berkeley.edu/SICI/>, standard koji osigurava dodjelu jedinstvenog identifikatora za dijelove serijske publikacije (broj, članak) bez obzira na medij (tiskani, elektronički, mikrooblik itd.).

¹³ Vidjeti NEDLIB, <http://www.bk.nl/nedlib>.

Literatura

- Bide, Mark. Standards for electronic publishing : an overview. The Hague : Koninklijke Bibliotheek, 2000.
- Butterfield, Kevin. Cataloger's and the creation of metadata systems : a collaborative vision at the University of Michigan. <http://www.oclc.org/oclc/man/colloq/butter.htm> (2000-07-11)
- Hakala, Juha. Metadata and identifiers for e-journals. // Digitising journals : Conference on future strategies for European libraries, Copenhagen, 13.-14. March 2000 : proceedings. Copenhagen : Danish National Library Authority; Danmark's Electronic Research Library, 2000. Str. 53-68.
- Langballe, Anne M. Hasund. Remote access electronic serials and the National Library of Norway. // International Cataloguing and Bibliographic Control 28, 3(1999), 78-79.
- Jones, Wayne. Dublin Core and serials : how well do the DC data elements describe e-serials? <http://www.mit.edu/people/waynej/dublincore.htm> (2000-06-29)

PRILOG 3

METAPODACI ELEKTRONIČKIH SERIJSKIH PUBLIKACIJA

Damir Pavelić

Ekonomski fakultet u Zagrebu, Knjižnica i dokumentacija

Uvod

Elektroničke serijske publikacije nalazimo u zbirkama digitalnih i hibridnih knjižnica, u bazama podataka njihovih nakladnika i raspačavatelja ili ih, što je u nas najčešći slučaj, nalazimo pojedinačno zastupljene i stavljene na WWW. Ovisno o ozbiljnosti i stručnosti načina predstavljanja elektroničke serijske publikacije, (ne)dostupan je i njen opis pomoću metapodataka.

Od knjižnica koje posjeduju baze podataka može se očekivati kvalitetan bibliografski opis. No, problem je u zastupljenosti upotrebe metapodataka jer danas, unatoč nastojanju stručne zajednice da se model opisa standardizira, ne postoji konsenzus o upotrebi jednog modela metapodataka.

Nakladnici i raspačavatelji u velikim bazama podataka koriste SGML (Standard Generalized Markup Language) pa je za očekivati da će kvalitetno opisati vlastite publikacije. Oni, međutim, nisu tako vični opisivanju izvora informacija kao knjižničari niti je SGML nastao u tu svrhu. Kako bi se uredilo mnoštvo podataka u takvim bazama podataka, potrebno ih je opisati nekim od modela metapodataka prilagođavajući ga korisničkim zajednicama.

Pojedinačno predstavljeni časopisi na WWW-u najčešći su slučaj u Hrvatskoj. Najveći je broj onih koji prate istoimeno tiskano izdanje. Pogledamo li koliko ih je opisano metapodacima, doći ćemo do poražavajućeg saznanja da samo 32,8% naslova sadrži neke metapodatke.¹ Budući da se ne koristi ni Dublin Core kao jedan od najpoznatijih modela kojeg podržava i W3C (World Wide Web Consortium), niti HTML tagovi kojim bi se mogli opisati *.html* dokumenti, otežano je pronalaženje serijskih publikacija. Razlog je tome što šira javnost nije upoznata s knjižničarskim i drugim modelima metapodataka, a stvaratelji i naručitelji stranica, sudeći po stanju stvari u nas, najčešće ne brinu o "nevidljivoj" dijelu stranice, odnosno o opisu onoga što se ne prikazuje na ekranu.

Razvoj portala, posebice strukovnih *gatewaya*, mogao bi olakšati pronalaženje željenih publikacija i osigurati kvalitetan metaopis. Iako takve usluge pokrivaju određeno područje i funkcioniraju kao svojevrsna baza podataka, svaka bi jedinica trebala biti opisana na standardiziran način kako bi pretraživanje bilo lakše, brže i jednostavnije.

¹ Vidjeti Prilog 2.

Upotreba metapodataka

Dva su osnovna načina upotrebe metapodataka: unutar dokumenta ili izvan njega u posebnoj datoteci. U oba slučaja metapodatke izrađuju katalogizatori ili osobe odgovorne za izradu web-stranica koje se tada smatraju vlasnicima opisanih dokumenata (časopisa, tekstova, stranica i sl.). Metapodaci u posebnoj datoteci mogu biti povezani s podacima koje opisuju tj. datoteka s metapodacima može se nalaziti u svojevrstnom katalogu metazapisa. Pri tome svaki metazapis ima izravnu vezu prema publikaciji ili tekstu. Takav se pristup primjenjuje na Manchesterskom sveučilištu. Pomoću zajednička kataloga i jedinstvena sučelja pretražuju se različite baze nakladnika, te na jedinstven i standardiziran način predočuju rezultati pretraživanja. Na isti način može se koristiti i UNIMARC-zapis. Sam zapis predstavlja metapodatke, a polje 856 sadrži hipervezu prema odgovarajućoj elektroničkoj serijskoj publikaciji.

Poseban je slučaj kada metapodatke izrađuje treća strana koja ne mora posjedovati ništa od onoga što opisuje. Tako, na primjer, portali i *gatewayi* često sakupljaju podatke o određenom predmetnom području, te ih na određeni način katalogiziraju i popraćuju opisom kako bi ih korisnici mogli lakše pronaći. Druga je mogućnost da stručnjaci u takvim servisima izrađuju metapodatke koje potom autori ili web-administratori jednostavno uvrstavaju u svoje .html dokumente. Tada su to također podaci izvan dokumenta, ali su posebnim servisima povezani s dokumentom što je od obostrane koristi i nakladniku/autoru i *gatewayu* koji te metapodatke izrađuje.

Metapodaci unutar dokumenta (.html, na primjer) nalaze se u čitatelju/korisniku interneta "nevidljivoj" sekciji tzv. zaglavlju (<HEAD>). One se ne vide na samoj web-stranici, ali služe nekim pretraživačima² da ga lakše pronađu.

Važno je napomenuti da metapodaci nisu nastali razvojem interneta (i kataložni su zapisi metapodaci). Ogromna količina informacija i publikacija na WWW-u samo je povećala potrebu lakšeg pronalaženja i pretraživanja te utvrđivanja vjerodostojnosti dokumenta. Od modela metapodataka najčešće se spominje Dublin Core Metadata Element Set (Dublinski osnovni skup elemenata metapodataka) koji je gotovo postao standard. Integracija s formatima RDF (Resource Description Framework) i XML (eXtended Markup Language) može znatno pogodovati širenju Dublin Corea i izvan knjižničarske zajednice. Tome pogoduje i prevođenje formata na jezike bibliografskih središta. Posljednji je slučaj prijevod Dublin Corea na poljski jezik.

Opis dokumenta elementima Dublin Corea vrlo je lako izraditi, čak i bez poznavanja elemenata opisa. Trenutno postoji nekoliko pomagala koji olakšavaju izradu opisa (zapisa), od kojih su neki automatizirani. Primjer je automatiziranog pomagala za izradu metapodataka iz .html dokumenta DC-dot (www.ukoln.co.uk/metadata/dcdot) (Slika 1). Unosi se samo odgovarajući URL, a program sam pregledava dokument i izrađuje zapis Dublin Corea u formatima .html ili RDF. Loša je strana programa što on, ako u zaglavlju doku-

² Neki pretraživači indeksiraju cijeli tekst dokumenta, neki samo dio, a neki uopće ne prolaze kroz <HEAD> dio pa tako niti upotreba metapodataka tj. odgovarajućih tagova ne utječe na rezultat pretraživanja. S druge strane, Metabrowser (www.metabrowser.spirit.net.au) pregledava zaglavlje tražeći metapodatke koje zatim može prikazati zajedno s web-stranicom.

menta ne postoje metatagovi, generira metapodatke a iz sadržaja stranice izabire one podatke koje smatra odgovarajućim za pojedino polje. Međutim, program nudi i ručno unošenje podataka u vrlo jednostavnom sučelju tako da se zapis može popraviti i na kraju integrirati u .html dokument ili katalog metazapisa.



Slika 1. Sučelje DC-dot programa za samostalan unos podataka

Samostalan upis podataka u odgovarajuća polja i potom gotov zapis nude i ovi programi:

- Reggie - metadata editor (www.metadata.net),
- TagGen Dublin Core edition (www.hissoftware.com)
- DC metadata template (www.lub.lu.se/cgi-bin/nmdc.pl)

S obzirom na postojanje nekoliko modela metapodataka javila se i potreba međusobne konverzije. Mnoge ustanove posjeduju velik broj MARC zapisa, a javlja se i potreba preuzimanja zapisa u drugim formatima pri čemu treba obaviti odgovarajuću konverziju. I na ovom području postoji nekoliko pomagala. Zbog velike zastupljenosti MARC formata jedan od nama najzanimljivijih jest onaj kojeg je razvio *Nordic Metadata Project*. On pretvara MARC (DANMARC, FINMARC ... nažalost ne i UNIMARC) u Dublin Core i nudi ispis u obliku običnog ili .html teksta. Pretvorba se obavlja unošenjem URL-a ili pak samostalnim upisom sadržaja. Više informacija o tom može se pronaći na adresi: <http://www.bibsys.no/meta/d2m/>.

Neki od alata nude i mogućnost usporedbe postojećih modela, razvoja vlastita modela metapodataka korištenjem već postojećih osnova ili pak razvoja potpuno novog modela metapodataka (Reggie, ROADS). Adrese na kojima se može više saznati o tome jesu:

- www.getty.edu
- www.ukoln.ac.uk/metadata/interoperability
- <http://purl.org/tools/indexhtml>

Praktičnu primjenu konverzije možemo pronaći u već spomenutom projektu Manchesterskog sveučilišta *Electronic Journals Application*. Da bi se moglo pristupiti časopisima dvaju dobavljača, bilo je potrebno prilagoditi format zapisa izdavača (SGML) standardiziranom formatu ispisa za Sveučilište tj. u kvalificirani Dublin Core zapis u XML formatu koji se pomoću programa *OmniMark* u toku prijenosa prema korisnikovu računalu prevodi u *.html* format i bez problema prikazuje na ekranu u standardnom web-sučelju. Tako izrađen zapis sadrži i veze prema cjelovitu tekstu (*.html* i *.pdf*) i prema ostalim bibliografskim jedinicama istog autora. U posljednjem je slučaju bilo potrebno razviti i dodatne lokalne elemente opisa koje kvalificirani Dublin Core ne posjeduje. Razvijena je, dakle, tzv. Shema koja je implementirana u XML zapis pored postojeće Dublin Core sheme, a dodatni su elementi, npr., *MUP.VID* tag za Volume ID, *MUP.PPF* za početnu stranicu i slično.

Model koji će određena tvrtka ili ustanova odabrati trebao bi, prije svega, odgovarati njezinim potrebama no često su nedostatne financije ograničavajući faktor. Zbog toga je potrebno surađivati sa sličnim ustanovama i probleme rješavati na razini konzorcija. Potrebno je odabrati i standardizirane modele ili one koji imaju izgled postati standardom (Dublin Core), te voditi računa o mogućnostima nadogradnje softvera. Kao i uvijek, treba naglasiti da najjeftinije rješenje nije i najbolje. Isto tako, posebnu pažnju treba posvetiti razmjeni podataka s ostalim ustanovama ako je to u planu poslovanja.

XML i RDF

XML nije prošireni oblik HTML jezika. Između njih postoji XHTML kao svojevrsan prijelazni oblik. HTML određuje strukturu podataka i način prikaza podataka na ekranu. Pomoću HTML-a određujemo gdje će što biti smješteno, kakve će se boje i tablice upotrebljavati, gdje će biti slike i sl. Ukratko, kako će sve što nudimo izgledati na ekranu korisnikova računala.

XML određuje isključivo strukturu podataka a podatke zapisuje u obliku obične tekstualne datoteke (*.txt*). Kao što mu naziv govori, proširiv je pa možemo prema potrebi definirati tagove na željenom jeziku. Pri tome treba voditi računa da je dokument valjan tj. da su svi upotrijebljeni tagovi napisani prema strogim pravilima XML-a. Uz pojam XML usko je povezan pojam DTD (Document Type Definition), poznat još iz SGML-a, te XSL (eXtensible Style Language). Oba su napisana na bazi XML-a. DTD se koristi za definiranje korištenih tagova i njihova sadržaja, dok se XSL koristi za prikaz tih podataka na ekranu. XSL je skriptni jezik koji se sastoji od niza tagova sličnih HTML-ovim i služi predstavljanju sadržaja XML datoteke na ekranu.

DTD i XSL mogu se nalaziti unutar XML datoteke ili izvan nje. Ako su smješteni unutar datoteke, onda je njihovo korištenje ograničeno samo na tu datoteku. Ako su u posebnoj datoteci, tada ih možemo višestruko koristiti za sve XML datoteke kojima takva definicija odgovara, odnosno pri čestoj obradi različitih podataka jednake strukture.

Iz proširivosti XML-a proizlazi nastanak različitih jezika za označivanje (Markup Languages), ovisno o potrebama određene struke. Jedan je od poznatijih samostalnih jezika za označivanje WML (Wireless Markup Language) napravljen za potrebe mobilne telefonije, odnosno WAP-usluge (Wireless Application Protocol) namijenjene pretraživanju WWW-a mobilnim telefonom. Postoji još i MathML (Mathematical ML) koji definira označivanje matematičkih oblika i sl.

Na temelju sintakse XML-a napravljen je i RDF (Resource Description Framework) koji služi kao okvir za definiranje i korištenje metapodataka. Ako usporedimo zapise u RDF i XML obliku, primijetit ćemo veliku sličnost – proizvoljni tagovi i jednostavan prikaz strukture podataka. Stoga je logično pitanje zašto se umjesto RDF-a odmah ne koristi XML.

XML se bavi strukturom podataka, ali ne propisuje način opisivanja podataka. Jednostavnu tvrdnju "Ime Prezime je autor <http://www.stranica.hr>" možemo napisati u XML-u na nekoliko načina koji će biti logički ispravni. No to ne osigurava pronalaženje traženih podataka i ispis rezultata. Ideja je RDF-a da osigura način prepoznavanja odnosa unutar grupe podataka. RDF funkcionira na tzv. načelu trojki, što je najjednostavniji model za opisivanje odnosa. Sastoji se od elemenata: izvor podataka - atribut - vrijednost. U navedenom slučaju trojka bi bila:

<http://www.stranica.hr> - DC.Creator - Ime Prezime

Pojedini autori kao članove trojke navode subjekt, predikat i objekt.

Ovakvo definiranje strukture i prepoznavanje odnosa trebalo bi, kad zaživi, omogućiti razmjenu podataka na jednostavniji način. Zamisao je o RDF-u i krenula od želje za jedinstvenim opisivanjem sadržaja WWW-a kako bi automatizirane službe i aplikacije tj. pretraživači mogli izdvojiti željeno. Iako je RDF još uvijek u fazi razvoja, podržala ga je zajednica Dublin Core³ i zajedno s W3C konzorcijem radi na njegovu razvoju pa je za očekivati rezultate u području koje zanima knjižničarsku struku. Valja napomenuti da na razvoju rade i druge neprofitne i komercijalne organizacije.

Već spomenuti DC-dot omogućuje i ispis u RDF formatu korištenjem DC elemenata opisa a postoje i pretvarači XML zapisa u MARC format i obratno. Te aplikacije u ovom radu nisu detaljnije obrađivane, no zainteresiranima su dostupne na adresama:

- www.logos.com/marc
- www.culture.fr/BiblioML

Zaključak

Kad govorimo o metapodacima, obično mislimo na opisivanje jedinica (knjiga, članaka, sadržaja općenito), odnosno često se ograničavamo na ono što je u početku i bila ideja Dublin Corea i drugih modela koji za opis koriste metapodatke ali, kako vidimo, stvari su se promijenile u smjeru razmjene podataka. Vidjeli smo da se uvelike radi i na standardizaciji

³ Svaka web-stranica na adresi Dublin Core Metadata Initiative ima i svoj RDF oblik uz korištenje elemenata Dublin Corea, vidjeti <http://purl.org/dc>.

određenih formata kako bi se opisani podaci mogli razmjenjivati zadržavajući pritom pojedinih potrebne za lokalnu obradu. Izrađuju se lokalne verzije Dublin Corea, prilagođene verzije XML-a, standardizira se RDF. Znatno doprinos u ovom području donosi razvoj XML-a kao jezika za opisivanje strukture podataka. Od velike nam koristi mogu biti mogućnosti kontrole i statistike koje krajnjim korisnicima nisu dostupne, a ustanovi mogu olakšati održavanje baze podataka i razmjenu podataka.

Poželjno je pratiti trendove razvoja i projekte knjižnica u inozemstvu koje znatno lakše dolaze do potrebnih poticaja za implementaciju softverskih i hardverskih novina. Samo nekoliko posjeta određenim web-stranicama bit će dovoljno da se uoči koja količina podataka postoji o temi koja je u nas slabo prisutna, kako u knjižničarskim krugovima tako i u informatičkim i poslovnim. Vrijeme je za primjenu u praksi, a početak korištenja Dublin Core metapodataka mogao bi znatno doprinijeti razvoju korištenja RDF-a i XML-a u hrvatskim knjižnicama i šire.

Literatura

<http://www.w3.org>
<http://www.ukoln.ac.uk>
<http://purl.org/dc>
<http://www.metadata.net>
<http://www.ifla.org>
<http://w3schools.com>
<http://www.xml.com>
<http://www.xml.org>
<http://www.ibm.com/developer/xml/>

Primjeri: Metapodaci teksta napravljeni pomoću DC-dot programa

Dublin Core

```
<link rel="schema.DC" href="http://purl.org/dc">
<meta name="DC.Title" content="Metapodaci elektroničkih serijskih publikacija">
<meta name="DC.Creator" content="Damir Pavelić">
<meta name="DC.Subject" content="metapodaci, elektroničke serijske publikacije,
XML, RDF, Arhivi, knjižnice, muzeji, AKM 2000">
<meta name="DC.Description" content="Tekst radionice održane na simpoziju Arhivi,
knjižnice, muzeji 2000, Rovinj, prosinac 2000. godine">
<meta name="DC.Publisher" content="Hrvatsko knjižničarsko društvo">
<meta name="DC.Contributor" content="Sofija Klarin, Sonja Pigac">
<meta name="DC.Date" content="22.01.2001.">
<meta name="DC.Type" scheme="DCMIType" content="Text">
<meta name="DC.Format" content="text / word .doc / 1.081.856 bajtova">
<meta name="DC.Identifier" content="http://jagor.srce.hr/akm">
<meta name="DC.Language" scheme="RFC1766" content="HR">
<meta name="DC.Rights" content="AKM 2000">
```

RDF

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF SYSTEM "http://purl.org/dc/schemas/dcmes-xml-20000714.dtd">

<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 about="http://jagor.srce.hr/akm">
    <dc:title>
      Metapodaci elektroničkih serijskih publikacija
    </dc:title>
    <dc:creator>
      Damir Pavelić
    </dc:creator>
    <dc:subject>
      metapodaci elektroničke serijske publikacije XML RDF
      Arhivi, knjižnice, muzeji AKM 2000
    </dc:subject>
    <dc:description>
      Tekst radionice održane na simpoziju Arhivi, knjižnice,
      muzeji 2000, Rovinj, prosinac 2000. godine
    </dc:description>
```

```
<dc:publisher>
  Hrvatsko knjižničarsko društvo
</dc:publisher>
<dc:contributor>
  Sofija Klarin, Sonja Pigac
</dc:contributor>
<dc:date>
  22.01.2001.
</dc:date>
<dc:type>
  Text
</dc:type>
<dc:format>
  text / word .doc / 1.081.856 bajtova
</dc:format>
<dc:language>
  HR
</dc:language>
<dc:rights>
  AKM 2000
</dc:rights>
</rdf:Description>
</rdf:RDF>
```