

INTEGRIRANJE METAPODATAKA U DIGITALIZIRANU SLIKU NA PRIMJERU DUBLIN CORE

INTEGRATING METADATA INTO A DIGITIZED IMAGE USING DUBLIN CORE

Laila Miletić-Vejzović*

Washington State University, Library, Pullman, SAD
e-mail: vejzovic@wsu.edu

Sažetak

U radu se opisuju postupci digitalizacije građe iz Zbirke fotografija Franka S. Matsure Knjižnice Državnog sveučilišta Washington, SAD: odabir građe za digitalizaciju, rješavanje autorskih prava, postupci skeniranja, istraživanje podataka o samoj građi, formalna i sadržajna obrada, te odabir elemenata opisa relevantnih za pretraživanje građe. Za izradu metapodataka odabran je format Dublin Core, a za bazu digitaliziranih fotografija program CONTENT Database.

Stavljanje digitalizirane zbirke fotografija F. S. Matsure na internet početak je nastojanja demokratizacije pristupa fondovima specijalnih zbirki Knjižnice Državnog sveučilišta Washington.

Ključne riječi: digitalizacija fotografija, Dublin Core

Summary

This paper describes: the procedures for the digitisation of photographs from the Frank S. Matura Photograph Collection, in the Library of Washington State University, USA; the selection of the material for digitisation; solving copyright problems; scanning procedures; research of the data on the material; formal and subject indexing; and selection of the descriptive elements relevant for searching the material. The Dublin Core format was chosen for metadata creation, and for the digitised photographs database of the Content Database program.

Keywords: digitisation of photographs, Dublin Core

Uvod

Neosporna je činjenica da su metapodaci izuzetno važni pri digitalizaciji fotografija, dokumenata, zemljopisnih karata, grafika itd. Sam postupak skeniranja vrlo je jednostavan i relativno jeftin. Potreban je dobar skener, računalno itd. i, naravno, osoba koja će taj fizički

* Research for this (activity, research, publication) was supported by a grant from the International Research & Exchanges Board, with funds provided by the US Department of State (Title VIII program) and the National Endowment for the Humanities. None of these organizations is responsible for the views expressed.

posao obaviti (prosječno 3 minute po slici na 300 dpija u TIFF formatu i 72 dpija u JPEG formatu). Rezultat su skenirane brojne slike pohranjene u digitalnom obliku. Kako bi digitalizirane fotografije postale funkcionalne tj. pretražljive i uporabljive potrebno ih je na odgovarajući način opisati i indeksirati. No, često se zanemaruje još jedna faza koju treba obaviti prije unošenja metapodatka. To je vrijeme istraživanja koje prethodi obradi (prosječno 15-20 minuta po slici).

Metapodaci

Pojam metapodatak vezan je uz fenomen interneta. To su podaci o podacima nastali iz potrebe da se ogromna količina informacija na internetu stavi pod neku vrstu nadzora odnosno "katalogizira" i postane dostupna pretraživačima. Prije informacijske revolucije, isključivo su knjižničari bili odgovorni za izradu bibliografskih zapisa. Posjedovali su potrebno profesionalno znanje i primjenjivali određene standarde i pravila. Informacijska revolucija nalagala je pronalaženje i primjenu novih i fleksibilnijih standarda koje bi mogli rabiti laici. Upravo je takav Dublin Core standard za metapodatke. Razvijen je u OCLC-u u Dublinu (Ohio). Njegov su razvoj potaknuli:

- jednostavnost – svi su elementi lako razumljivi i ne-katalogizatorima, kao i različitim predmetnim stručnjacima;
- semantička interoperabilnost - na različite načine opisani modeli onemogućuju pretraživanje različitih disciplina na internetu; uporabom pristupnica koje svi mogu razumjeti i koje ujedinjuju standardi za opis sadržaja, povećava se mogućnost semantičke interoperabilnosti različitih disciplina;
- međunarodni konsenzus – zbog ogromne količine građe na webu koja nije indeksirana pa stoga ni pretražljiva, postignut je međunarodni konsenzus o potrebi stvaranja efikasne infrastrukture za pronalaženje informacija; ta je činjenica bila važna za uspjeh Dublin Corea i njegovu prihvaćenost diljem svijeta od Sjeverne Amerike i Europe, do Australije i Azije;
- ekstenzibilnost – Dublin Core ekonomičnija je alternativa od složenijih modela za pristupnice, kao što je, na primjer, potpuno katalogiziranje u MARC formatu u knjižnicama;
- uz fleksibilnost i ekstenzibilnost, za kodiranje strukture Dublin Core uključuje i složeniju semantiku unutar bogatijih opisnih standarda;
- modularnost metapodataka na webu – potreba za različitim vrstama metapodataka na webu iziskuje i infrastrukturu koja podržava koegzistenciju samostalno proizvedenih i održavanih metapodataka (paketa) koji se međusobno nadopunjuju; Konzorcij za World WideWeb (World Wide Web Consortium: W3C) implementirao je arhitekturu za metapodatke za web. Okvir za opis građe (Resource Description Framework: RDF) oblikovan je tako da može zadovoljiti razne potrebe za metapodatke.

Dublin Core skup je znakova i njihovih značenja, kojim se opisuje bilo koji digitalan skup podataka. Na prvi je pogled knjižničarima izgledalo da Dublin Core sadrži dosta elemenata koji se primjenjuju u MARC formatu. Međutim, elementi Dublin Corea stavljeni

su u novi, razumljiviji kalup koji se može jednostavnije primijeniti u mnogobrojnim prigodama. Standard ne iziskuje kontroliran rječnik jer se smatralo da treba biti prilagodljiv potrebama opisa dokumenata koji pripadaju najrazličitijim područjima. Isto tako, Dublin Core ne definira sheme za indeksiranje, kodiranje, programe, izgled itd. Elementi su ponovljivi pa se konačan zapis razlikuje prema potrebama i sadržaju digitalnih podataka.

Petnaest je osnovnih elemenata metapodataka Dublin Core: Naslov, Stvaratelj, Tema, Opis, Nakladnik, Suradnik, Datum, Tip, Format, Identifikator, Izvor, Jezik, Odnos, Obuhvat i Vlasnička prava. Svaki se od navedenih elemenata može indeksirati, pretražljiv je i individualno se može točno definirati odnosno prevesti u drugi format (*mapping*). Kao konkretan primjer osvrnut ću se na nedavno završen projekt digitalizacije jedne od naših fotografskih zbirki.

Iniciranje digitalizacije fotografija u državi Washington – natječaj za pilotno testiranje

U jesen 1999. godine Državna knjižnica Washington raspisala je natječaj za različite projekte digitalizacije. Glavni je cilj ovog natječaja bio testiranje različitih praksi kako bi se utvrdila najbolja i izradile upute za buduće digitalne projekte. Knjižnica *Manuscripts, Archives & Special Collections* prijavila se s prijedlogom digitaliziranja jedne od svojih fotografskih zbirki.

Naš je prijedlog i projekt bio prihvaćen, kao i zatražena financijska sredstva. Svim ustanovama-sudionicama u tom testiranju uvjetovana je primjena istog programa za digitalnu bazu podataka, tzv. CONTENT Database. Ta je digitalna baza podataka razvijena na Sveučilištu Washington. CONTENT software suite version 2.2 sastoji se od aplikacije za poslužitelj, stanica za prihvaćanje podataka te JAVA i HTML pretraživačkog klijenta/korisnika. Drugi je uvjet bio da sve ustanove primijene isti standard Dublin Core za integriranje metapodataka za digitalne fotografije, rukopise, tiskane tekstove i novinske članke. Svaka ustanova, međutim, mogla je izraditi *mapping*, indeksiranje i prikaz podataka prema svojstvima građe i potrebama korisnika.

Mi smo odlučili digitalizirati zbirku fotografija Franka S. Matsure. Za tu smo se zbirku odlučili zbog više razloga. Pri izboru vrlo je značajan bio sadržaj zbirke budući da Matsurine fotografije pružaju jedinstven pregled lokalne povijesti pokrajine Okanoga. No najvažniji je svakako razlog što smo već imali detaljni inventar i opis ove zbirke. Zbirka je, dakle, već bila obrađena i brojila 1241 sliku. Za četveromjesečno razdoblje trajanja projekta predložili smo da se potpuno obradi 500-600 fotografija (skeniranje, integracija metapodataka i indeksiranje). Zbog tipičnih birokratskih razloga, tj. raznih ugovora i financijskih sporazuma, radom na projektu započeli smo tek krajem travnja 2000., pa je projekt trajao do kraja kolovoza 2000. Na kraju smo ipak uspjeli realizirati potpunu digitalizaciju i indeksiranje za cijelu zbirku od 1214 fotografija.

Za navedenu smo se zbirku odlučili i zbog toga što smo smatrali da za nju imamo vlasnička i autorska prava. No, ispostavilo se da još jedna ustanova posjeduje zbirku Matsurinih fotografija kao i negative nekih naših fotografija. Kako je bilo nemoguće točno ustanoviti autorska prava, osobno sam dva mjeseca pregovarala za dobivanje pismenog dopuštenja za "objavljivanje na internetu". Stečeno iskustvo dragocjeno je za buduće projekte.

U toku priprema za projekt dosta je vremena iziskivala i izrada troškovnika tj. potraživanja sredstava od Knjižnice države Washington.

Jedan je od glavnih ciljeva projekta bio da se preko web-stranice korisnicima širom države Washington, kao i svim ostalim korisnicima interneta, omogući pristup svakoj pojedinoj fotografiji iz Matsurine zbirke. Smatrali smo da će testiranje pomoći i pri planiranju budućih projekata, te dati uvid u troškove i pitanja logistike pri prijenosu fotografija u digitalan oblik. Također smo željeli ustanoviti koliko vremena iziskuje poduka osoblja.

Na kraju su se svi naši ciljevi ostvarili. Skenirana je i obrađena metapodacima 1241 fotografija, te je omogućen pristup zbirci (Frank S. Matura Image Collection) preko interneta na adresi <http://www.wsulibs.wsu.edu/holland/masc/xmatura.html>. No, što je još važnije, sada imamo jasnu spoznaju o tome koliko vremena i troškova iziskuje organizacija i realizacija digitalnih projekata.

Za pilotni projekt napravili smo dvije baze podataka, jednu s 1167 fotografija i drugu, tzv. *compound document database* sa 74 fotografije, ukupno 1241 fotografija. Također smo dizajnirali korisničko sučelje za obje baze podataka. Uz podatke o fotografu i zbirci, upute o pretraživanju baze podataka i opis projekta, ono nudi brojne mogućnosti pretraživanja (opće pretraživanje kroz sva polja i niz unaprijed definiranih pristupnica). *Compound document database* sadrži 74 stranice albuma i dopušta korisniku pregledavanje stranice po stranicama.

Skeniranje

Postupak skeniranja bio je relativno brz, čak puno brži od prethodnih skeniranja budući da smo upotrijebili novi skener Microtek 9600XL. Fotografije su skenirane u skupinama od 3 do 6 (ovisno o veličini) na desktopu, i istovremeno prenošene u program Photoshop. Ovaj je korak uštedio dosta vremena jer nije bilo potrebno ponovno otvaranje i pregledavanje svake pojedinačne fotografije.

Još se jednom potvrdilo moje stanovište da je sam postupak skeniranja relativno jeftin i brz. Najskuplje je i najdugotrajnije bilo proučavanje fotografija i prikupljanje podataka za kvalitetan opis. Studentica knjižničarstva, koja je bila posebno stipendirana za te poslove, u preliminarnom je istraživanju zbirke provela oko 330 sati. Budući da je zbirka fotografija usko vezana uz pokrajinu Okanoga, studentica je pročitala na mikrofilmu sve lokalne novine iz razdoblja u kojem se Matura aktivno bavio fotografijom tj. od 1907. do 1913. Također je pročitala brojne knjige i konzultirala druge zbirke u našem odjelu. Nakon opsežnih priprema, započela je obrada. Za unosenje metapodataka, reviziju naslova, dodavanje zapisa, predmeta i žanra te indeksiranje potrošeno je oko 30 minuta po fotografiji. U tih 30 minuta uključeno je i vrijeme provedeno na općoj reviziji.

Iako je odabrana zbirka bila potpuno obrađena s dobrim pristupom na zbirnoj razini, ustanovljeno je da pojedinačne fotografije nisu bile adekvatno obrađene. Neke od njih bile su pogrešno ili loše opisane, a o nekima nije bilo nikakvih metapodataka. Matura nije navodio čak ni datum na poledini pojedinih fotografija.

Edukacija osoblja za upotrebu programa CONTENT database i raznih administrativnih funkcija nije iziskivala puno vremena. Dobro nam je poslužio *Thesaurus for graphic mate-*

rials za predmete i žanr te predmetne odrednice Kongresne knjižnice (Library of Congress Subject Headings). Iako Dublin Core ne iziskuje primjenu kontroliranog rječnika, ipak smo, kad je bilo moguće, primjenjivali navedene standarde zbog lakše razmjene podataka i zadržavanja konzistencije cjelokupnog kataloga Knjižnice.

Izradili smo upute za primjenu Dublin Corea, razne specifikacije i *mapping*. Kod *mappinga* uvijek je važno voditi računa o vrsti građe određene zbirke kao i mogućim korisnicima i namjeni. Stoga je osnovno načelo bilo da se unesu oni podaci koji će najviše koristiti pri pretraživanju i pronalaženju građe na mreži. Stvaranju metapodataka posvetilo se dosta vremena i pažnje.

Na primjer, prvi je element *Naslov* kao i u Dublin Coreu. Odlučili smo da to treba biti pretražljivo i obvezno polje. U napomeni smo naveli da polje može uključiti natpis koji se nalazi na samoj fotografiji, da se datum stavlja u okrugle zagrade i da treba upotrebljavati velika slova za sve imenice u naslovu. Za Dublin Core element *Stvaratelj* odlučili smo upotrijebiti oznaku *Fotograf*. Taj je element podatka također pretražljiv, a autorizirani je oblik imena Matsura, Frank S. određen prema normativnoj datoteci za vlastita imena Kongresne knjižnice (Library of Congress Name Authority File).

Zaključak

Pilotni projekt bio je od velike koristi pri stjecanju iskustva u digitalizaciji općenito, a naročito pri digitalizaciji fotografija. Uspjeli smo ustanoviti i razviti najbolje prakse i smjernice za buduće digitalne projekte. Također, stekli smo i dobro iskustvo u vezi *mappinga* i već smo neka polja modificirali za buduće digitalne projekte. Potvrđeno je da je skeniranje relativno jeftin postupak, a integriranje metapodataka u digitalnu sliku najskuplje i iziskuje najviše vremena. Nadalje, mnogo je vremena bilo potrošeno na istraživanje koje je prethodilo unošenju metapodataka na primjeru Dublin Corea. Važno je napomenuti da smo svaku fotografiju pokušali opisati što podrobnije pri prvom unosu podataka jer je to jeftinije i ekonomičnije od vraćanja fotografijama i naknadnih preinaka radi usavršavanja i poboljšavanja pristupnica, opisa itd. Moto je, dakle, bio da se prvi put izradi dobar opis i da se ne vraća pojedinim fotografijama.

Na kraju bih htjela napomenuti još jednu vrlo važnu činjenicu. Nova tehnologija i postojeći standardi omogućili su demokratizaciju pristupa informaciji. To osobito vrijedi za knjižnice koje posjeduju stare i rijetke zbirke. Unatoč otvorenom necenzuriranom pristupu, koji je oduvijek bio pravilo u našim knjižnicama, ipak je naša građa bila dostupna samo korisnicima koji su fizički morali doći kako bi ustanovili što točno posjedujemo. Sada s digitalizacijom i objavljivanjem građe na internetu korisnici mogu iz udaljenih lokacija doznati i utvrditi što knjižnice posjeduju prije samog dolaska što im štedi vrijeme i novac. To nama knjižničarima predstavlja veliku satisfakciju.

Literatura

- Anon. Metadata and the Dublin Core : Emerging standards for digitization projects Information Intelligence. // Online libraries, and microcomputers 16, 12(December 1998), 1-5.
- Bearman, David et al. A common model to support interoperable metadata : progress report on reconciling metadata requirements from the Dublin Core and INDECS/DOI Communities. // D-lib magazine 5.1 Jan. 1999. <http://www.dlib.org/dlib/january99/bearman/01bearman.html> (2000-11-06).
- Chepesiuk, Ron. Organizing the Internet : the "core" of the challenge. // American libraries 30, 1(1999), 60-63.
- Contextual database. <http://www.wsu.edu/~cornish/testbed.html>
- Dublin Core Metadata Initiative. <http://purl.org/DC/education/index.htm#18>
- Forsberg, Kerstin; Lars Dannstedt. Extensible use of RDF in a business context. // Computer networks 33, 1-6(2000), 347-364.
- Frank S. Matsura Image Collection. <http://www.wsulibs.wsu.edu/holland/masc/xmatsura.html>
- Gilliland-Swetland, Anne J.; Yasmin B. Kafai; William E. Landis. Application of Dublin Core metadata in the description of digital primary sources in elementary school classrooms. // Journal of the American Society for Information Science 51, 2(2000), 193-201.
- Holland, Matt. Discovering online resources across the humanities : a practical implementation of the Dublin Core. // The electronic library 16, 6(1998), 407-408.
- Hunter, Jane; Liz Armstrong. A comparison of schemas for video metadata representation. // Computer networks 31, 11-16(1999), 1431-1451.
- Medeiros, Norm. Making room for MARC in a Dublin Core world online. // Wilton 23, 6(1999), 57-60.
- Oder, Norman. Cataloging the net : can we do it? // Library journal 123, 16(1998), 47-49.
- Resource Description Framework (RDF). <http://www.w3.org/RDF/>
- Sutton, Stuart A. Conceptual design and deployment of a metadata framework for educational resources on the Internet. // Journal of the American Society for Information Science 50, 13(1999), 1182-1192.
- Sutton, Stuart A. et al. Networked information discovery and retrieval for educational materials on the Internet : metadata development, deployment, and evolution. // Journal of the American Society for Information Science 35(1998), 54-62.
- WATCH : Writers, Artists, and Their Copyright Holders. <http://www.hrc.utexas.edu/watch/watch.html>
- Weibel, Stuart et al. Metadata preconference. // The serials librarian 38, 1-2(2000), 5-14.
- Weibel, Stuart. The state of the Dublin Core Metadata Initiative. // D-lib magazine 5 (4 Apr. 1999) <http://www.dlib.org/dlib/april99/04weibel.html> (2000-11-06).
- Weibel, Stuart. The state of the Dublin Core Metadata Initiative. // Bulletin of the American Society for Information Science 25, 5(1999), 18-22.