

METAPODACI U ORGANIZACIJI PODATAKA O ELEKTRONIČKOJ GRAĐI

MIRNA WILLER

Nacionalna i sveučilišna knjižnica, Zagreb

Informacije o metapodacima, njihovu razvoju, opisima njihovih zadaća i upotrebe povećavaju se gotovo svakodnevno. Stoga je potrebno uočiti jednu "čvrstu točku", izvor bibliografskih podataka koji će omogućiti daljnje praćenje ovoga područja. Takav se izvor najčešće prepoznaje u bibliografiji na IFLANET-u, IFLA-inoj mrežnoj stranici: <http://www.ukoln.ac.uk/II/metadata>.

Drugi je vrijedan izvor podataka mrežna stranica UKOLN-a (*UK Office for Library and Information Networking, University of Bath*): <http://www.ifla.org/II/metadata.htm>

DEFINICIJE: KATALOGIZACIJA NASUPROT STVARANJU METAPODATAKA

U predavanju pod naslovom "Cataloguing vs. Metadata: Old Wine in New Bottles?", održanom na 64. konferenciji IFLA-e u Amsterdamu u kolovozu 1998., Stefan Gradmann razmatrao je različite definicije metapodataka. Osnovna je nakana Gradmannova prikaza upozoriti na neodrživost mišljenja da su stvaranje metapodataka i "konvencionalno" katalogiziranje do određene mjere komplementarni postupci. Upravo suprotno, oni su temeljno različiti, pa čak proturječni radni modeli, s time da se bitno razlikuju i radni koncepti na kojima se temelje oba modela.

Gradmann najprije citira definiciju metapodataka Rachel Heery:

"Dobro poznati knjižnični kataložni zapis može se opisati kao metapodatak pod pretpostavkom da je kataložni zapis 'podatak o podacima'."

Neprimjerenost te definicije Gradmann nalazi u tome što ona ne upozorava na potrebu identificiranja nekih problematičnih mjesta koje pred knjižničare stavljaju metapodaci ili koje će oni u budućnosti prouzročiti. Ako se prihvati takav stav, knjižničari će naprosto katalogiziranje poistovjetiti s posebnom vrstom aktivnosti stvaranja metapodataka i zanemariti sve složene neknjižnične postupke vezane uz tu riječ.

Dvije definicije koje slijede opisuju metapodatke, u ovom slučaju Dublin Core, na isti način, nekom vrstom pojednostavnjena katalognog formata. Takvo je promatranje tipičan rezultat prilagodljivosti DC-a i već definiranih postupaka konverzija između MARC formata i DC-a.

“Metapodatak je podatak o podacima, i stoga pruža osnovne informacije kao što su ime autora, naslov djela, datum nastanka, odnosi prema srodnim djelima itd. Jedan je prepoznatljiv oblik metapodataka katalog na listićima u knjižnici; informacija je na tom listiću metapodatak o knjizi. Možda i ne znate, vi se svakoga dana služite metapodacima u svakodnevnom poslu...” (Paul Miller).

“Metapodatak nije doista ništa drugo do podatak o podacima; upravo to je i TEI header [oznaka zapisa, adresa] ili bilo koji drugi oblik opisa. Mi to zovemo katalogiziranjem, ali za neke taj je termin prepun prekomjerne prtljage, kao što su Angloamerička katalogna pravila i USMARC. Na neki način to je situacija za koju se može reći: “nije šija, nego vrat”, ali metapodatak je dobar neutralan termin koji pokriva sve temelje” (Priscilla Caplan).

Bitno je različita definicija Berners-Leea: dok se opći zadaci mogu poistovjetiti s katalognom aktivnošću, kontekst korištenja informacija (softverski agenti a ne knjižnični korisnici) i izričita briga za efikasnost podrazumijevaju postupke (brzina i lakoća) koji nisu karakteristični za katalognu obradu.

“Metapodatak je strojnorazumljiva informacija o mrežnoj građi ili o drugome... Izraz ‘strojnorazumljiva’ je ključ. Govorimo o informacijama kojima se služe softverski agenti kako bi nam olakšali život, osigurali da se držimo načela, zakona, provjeravali možemo li vjerovati u ono što radimo te činili da se sve događa jednostavnije i brže.”

Razlika postaje još jasnijom kada se uzme u obzir drugi aspekt koji je potaknuo nastajanje DC-a, prema riječima Stuarta Weibla:

“Jedan od prvobitnih motiva za niz Dublin Core radionica bila je zamisao da sami autori mogu izraditi svoje vlastite opise.”

Dakle, metapodatak ne izrađuje profesionalni katalogizator, te se proces obrade stoga razlikuje. Uz to, prvobitno je u žarištu DC radionica bilo “olakšati otkrivanje građe u mrežnom okruženju” (Carl Lagoze) pa metapodatak služi za otkrivanje/pronalaženje elektroničke građe, a ne prije svega njezinu opis. Navedeni pristup metapodacima pokazuje da oni samo usputno pripadaju opisnoj paradigmi knjižničnoga katalogiziranja.

Potpunija je definicija, rezimira Gradmann, koja sadrži implicitne i eksplicitne pretpostavke vezane uz pojam metapodatak: metapodaci su namijenjeni za *korištenje u kontekstu* drukčijem od knjižničnih kataloga, karakteristično je da *ih ne stvaraju profesionalni katalogizatori*, zamišljeno

je da se *izrađuju efikasnije od kataložnih zapisa*, pokrivaju *posebnu vrstu građe (elektroničku građu)* i *odnos između metapodataka i građe na koju se odnose bitno*, je različit od odnosa između kataložnog zapisa i knjige koju posjeduje knjižnica.

Bitna razlika u navedenom odnosu postaje izričito uočljiva kada se usporedi odnos između bibliografskog zapisa i knjige s odnosom između metapodataka i elektroničke građe. U knjižničnim sustavima bibliografske zapise nadopunjuju zapisi o fondu koji sadrže "pokazatelje" (mjesna signatura) koji upućuju na mjesto na kojemu se nalazi knjiga. Često je dodatno potrebno vrijeme knjižničnog osoblja koje opskrbljava korisnika knjigom. Osnovno na što ovaj opis postupka dolazanja do knjige želi upozoriti jest da kontekst upotrebe "tradicionalne" građe ima male ili gotovo nikakve posljedice na bibliografski zapis i kataložne postupke. Zapis metapodataka, nasuprot tome, sadrži detaljan opis o načinu pristupa dokumentu kao i njegovu mrežnu adresu, što omogućava izravnu dostupnost dokumenta. Metapodaci su dio određene tehničke informacijske infrastrukture: stvarna vrijednost zapisa metapodataka određena je u velikoj mjeri činjenicom da pristupni pokazatelji sadržani u zapisu funkcioniraju te da zadovoljavaju tehničke zahtjeve aplikativnog softvera koji se koristi za pristup informacijama. Upravo zato stručnjaci koji se bave metapodacima vrlo ozbiljno razmatraju problem "prekinutih veza" i razvoj sustava imenovanja elektroničkih dokumenata: URL (*Uniform Resource Locator*), PURL (*Persistent URL*) i URN (*Uniform Resource Name*).

Dostupnost elektroničkom dokumentu, dakle zapis metapodatka, vezan je, nadalje, uz brz razvoj Internetskih standarda, ali i legislative koja štiti i daje prava na korištenje.

PROJEKT DESIRE: VRSTE METAPODATAKA

Projekt DESIRE koji je započeo kao projekt Europske komisije DGXIII DESIRE: *Development of a European Service for Information on Research and Education*, nastavlja istraživanja i njihovu primjenu kao zasebna organizacija (<http://www.desire.org>)

Ovaj projekt u središte zanimanja postavlja metapodatke za široko shvaćene "informacijske izvore", odnosno mrežnu (udaljenu) elektroničku građu, definirajući ih kao podatke koji opisuju attribute građe. Karakteristično je za metapodatke da podržavaju različite zadaće, poput lociranja, otkrivanja, dokumentiranja, procjenjivanja, odabira i drugo. Te zadaće može obavljati čovjek kao krajnji korisnik ili njegovi (ljudski ili automatizirani) posrednici.

Formalnija je definicija metapodataka da je to "podatak pripisan objektima tako da potencijalni korisnici ne trebaju temeljito poznavati metapodatak ili njegove karakteristike." To znači da će u nepreglednu prostoru Interneta, efikasno upravljanje mrežnim informacijama sve više ovisiti o efikasnom upravljanju metapodacima. Stoga, metapodaci nisu samo ključ za otkrivanje, već će oni postati temeljnima za efikasno korištenje pronađene građe (uspostavljanjem tehničkih

ili poslovnih okruženja u kojima se mogu koristiti), kao i za interoperabilnost između protokolarnih područja djelatnosti.

Poslužit ćemo se nadalje općeprihvaćenom tipologijom formata metapodataka koja je izrađena za potrebe ovoga projekta (<http://www.ukoln.ac.uk/metadata/DESIRE/overview>). Procijenjeno je da jedan monolitni format metapodataka neće biti univerzalno prihvaćen. Razlog tome nisu samo tehnička i stručna ulaganja u pojedine formate. Različite su djelatnosti uložile velike napore u razvoj specijaliziranih struktura za detaljan opis građe kojim se bavi dotična djelatnost te razvile odgovarajuće informacijske sustave i visokospecijalizirana znanja koja se ulažu u razvoj i održavanje formata.

Formati su svrstani u tri skupine koje odražavaju spektar metapodataka od jednostavnih do sve složenijih u odnosu na potpunost i strukturu, i koje, ujedno, upućuju na zajednička svojstva pojedinih formata unutar skupine, njihove sličnosti i razlike. Naravno, granice između opisanih skupina nisu čvrste.

	Prva skupina	Druga skupina	Treća skupina
Svojstva zapisa	Jednostavni formati Vlasnički Indeksiranje teksta	Strukturirani formati Standardi u nastajanju Struktura polja	Bogati formati Međunarodni standardi Složena struktura polja
Formati zapisa	Lycos Altavista Yahoo itd.	Dublin Core IAFA uzorci RFC 1807 SOIF LDIF	ICPRS CIMI EAD TEI MARC

Prva skupina uključuje uglavnom nestrukturirane podatke, obično automatski izvučene iz građe (potpun tekst) i indeksirane za pretraživanje. Podaci nemaju izričito značenje i ne podržavaju pretraživanje po poljima podataka. Uglavnom ih stvaraju mrežni pretraživači (*crawlers*) koji su manje uspješni budući da su po opsegu globalni, zahvaćaju ograničene opise i nisu selektivni. Upravo zbog tih razloga, takve usluge nastoje razviti metapodatke koje autori uključuju u svoje mrežne stranice kako bi bile pretražive (tj. sakupljene od pretraživača). U okviru tih usluga razmatra se zajednički prikaz za razmjenu takvih metapodataka između globalnih kazala i drugih usluga, i time sakupljanje potpunijih metapodataka.

Druga skupina uključuje podatke koji sadrže dovoljno informacija da korisnik može procijeniti moguću vrijednost ili važnost građe a da je ne treba pregledati ili se na nju povezati. Podaci su strukturirani i podržavaju pretraživanje po poljima podataka, a zapisi su, uglavnom, dovoljno

jednostavni da ih može izraditi nestručnjak-korisnik. Ti su opisi uglavnom zasebni objekti i ne sadrže različite odnose između objekata. Oni su uglavnom, ali ne nužno, ručno izrađeni ili ručno doradjeni automatski izrađeni opisi i sadrže različite opisne i druge attribute. Izrađeni su tako da se izravno mogu uvrstiti u kazala usluga za pronalaženje ili da su pretraživi (sakupljeni: *harvested*).

Usluge u ovom području uključuju OCLC-ov NetFirst koji koristi vlastiti interni format i *Electronic Libraries Programme* u Velikoj Britaniji. Navedene usluge selektivno opisuju građu pa su stoga prilično skupe.

Treća skupina sadrži formate s potpunijim opisima koji mogu služiti pri lociranju i otkrivanju građe, ali oni imaju i zadaću pri dokumentiranju objekata, pa i skupina objekata. Njihov je nastanak povezan sa znanstvenim aktivnostima, zahtijevaju specijalizirano znanje pri stvaranju i održavanju, a brinu se za specijalizirane zahtjeve određene djelatnosti. Formati u ovoj skupini posebno su izražajni, tj. omogućuju hvatanje različitih odnosa između objekata na različitim razinama.

Formati su ove skupine ICPRS: *Inter-university Consortium for Political and Social Research* za opis skupova podataka za društvene znanosti, EAD: *Encoding Archive Description* za opis arhivskoga gradiva, CIMI: *Computer Interchange of Museum Information* za opis muzejske građe, TEI: *Text Encoding Initiative* - svi su navedeni formati verzije (DTD: *Data Type Definition*) formata SGML: *Standard Generalized Mark-up Language*. Unutar skupine formata knjižnične zajednice to su MARC: *Machine Readable Cataloguing* formati od kojih razlikujemo UNIMARC, MARC 21 (harmonizirani USMARC i CAN/MARC za bibliografske podatke), UKMARC, INTERMARC itd. te analogne formate knjižnične zajednice kao Pica+ (format nizozemskog sustava Pica) i MAB (njemački nacionalni format).

Zajednice koje upotrebljavaju bogate formate, pokazuju sve očitiju potrebu da informacije o svojoj građi otvore širem spektru korisnika. Kako će se to postići, nije još sasvim poznato. Jedna je mogućnost eksportiranje "pretraživog dijela" zapisa u druge sustave. Takva nastojanja daju navijestiti da formati srednje skupine postaju mnogo važniji kao pristupni put opće namjene, vjerojatno s vezama prema bogatijim formatima odnosno, gdje je to potrebno, zapisima specifičnim za djelatnost. Međutim, treba spomenuti da primjena protokola za povezivanje otvorenih sustava Z39.50 (ANSI/NISO Z39.50-1995 *Information Retrieval Application Service Definition and Protocol Specification for Open Systems*) uključuje pretraživanje građe temeljene na MARC-u, tj. formatu ISO 2709. Ta se slika postupno mijenja kako se uvode drugi profili i drugi formati kao npr. RDF (*Resource Description Framework*).

U zaključku pregleda metapodataka projekta DESIRE kaže se da je očito da organizacijska, uslužna i tehnološka okruženja usluga koje se bave otkrivanjem građe, nisu stabilna te da nije moguć siguran odabir pristupa ili pouzdano predviđanje budućih scenarija. Značajno je,

također, da ne postoji usluga koja bi se bavila tim razvojem. Pojedinačni interesi, natjecateljske prednosti, integracija s legalnim sustavima ili navike i iskustva, uvijek će značiti da postoje razlike u pristupu.

BILJEŠKE

Berners-Lee, Tim. Authentication of metadata. <http://meta2@mrrl.lut.ac.uk> (1998-01-23).

Caplan, Priscilla. You call it corn, we call it syntax-independent metadata for document-like objects. <http://info.lib.uh.edu/pr/v6/n4/caplan4.html>

Gradmann, Stefan. Cataloguing vs. metadata: old wine in new bottles? <http://www.ifla.org/IV/ifla64/007-126e.htm>

Heery, Rachel. Metadata formats. <http://www.ukoln.ac.uk/BIBLINK/wp1/d1.1>

Lagoze, Carl. From static to dynamic surrogates. Resource discovery in the digital age. <http://www.dlib.org/dlib/june97/06lagoze.html>

Miller, Paul. Metadata for the masses. <http://www.ariadne.ac.uk/issue5/metadata-masses>

Weibel, Stuart. Re: authentication of metadata. <http://meta2@mrrl.lut.ac.uk> (1998-01-23)