

INTEROPERABILNOST SADRŽAJA METAPODATAKA

INTEROPERABILITY OF METADATA CONTENT

Mirna Willer

Nacionalna i sveučilišna knjižnica, Zagreb
e-mail: mwiller@nsk.hr

Sažetak

Rad opisuje razine interoperabilnosti informacijskih sustava u arhivima, knjižnicama i muzejima koje je potrebno realizirati u izgradnji sustava i kataloga/obavijesnih pomagala kako bi se korisniku omogućila usluga pristupa metapodacima o građi bez obzira u kojoj je od tri navedene ustanove ona pohranjena. Naglašena je važnost interoperabilnosti sadržaja metapodataka.

Ključne riječi: arhivi, knjižnice, muzeji; interoperabilnost sadržaja metapodataka; Z39.50; ISO 23950; Bath profil; MARC formati; Dublin Core; ISO 2709; SGML; XML; RDF

Summary

The article describes levels of interoperability between information systems in archives, libraries and museums which are necessary to be realized in the system and catalogue/finding aid design in order to provide the user with services that will allow him to retrieve metadata about the material regardless of which of the three mentioned types of institutions have it in their custody. Relevance of the interoperability of metadata content is being stressed.

Keywords: archives, libraries, museums; interoperability of metadata content; Z39.50; ISO 23950; Bath profile; MARC formats; Dublin Core; ISO 2709; SGML; XML; RDF

Uvod

Ideja je seminara *Arhivi, knjižnice, muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture* razmijeniti znanja, iskustva i način rada između triju djelatnosti o postupcima organizacije informacija o građi za koju brinu, pronaći razine djelovanja na kojima se mogu koristiti pomagala druge djelatnosti i razine koje se zajednički mogu razvijati bez obzira na specifičnosti građe, te, po mogućnosti, razviti model usluge, odnosno takvog sustava koji će omogućiti korisniku bilo koje od tri spomenute ustanove ili potencijalnom korisniku koji "plovi" internetom da dobije jedinstven odgovor na svoj upit bez obzira gdje je građa pohranjena. Zato je i potrebno ispitati razine "komunikativnosti" informacijskih sustava – njihovu interoperabilnost, da bismo mogli odgovoriti na pitanje jesu li sadržaji metapodataka interoperabilni – što je krajnji rezultat sustava u obliku odgovora na korisnički upit.

Definicije

Pojam interoperabilnost/interoperabilan, gotovo sve-prisutan pojam, različito se definira. Slijedi nekoliko definicija koje su nastale u jednom širem kontekstu u koji, dakako, spadaju i teme kojima se bave sudionici seminara *Arhivi, knjižnice, muzeji*.

Interoperabilan

onaj koji može zajedno djelovati

(*Concise Oxford Dictionary*, 9th edition)

To je najopćenitija definicija.

Interoperabilnost

jest mogućnost sustava ili proizvoda da radi s drugim sustavima ili proizvodima bez da korisnik ulaže u to poseban napor. Interoperabilnost postaje kvaliteta povećane važnosti za proizvode informacijske tehnologije u ostvarenju koncepta "mreža je računalo". U tom se smislu termin mnogo koristi pri marketinškom opisu proizvoda.

(whatis.com)

Interoperabilnost

jest mogućnost dvaju ili više sustava ili komponenti da razmjenjuju informacije i da koriste informacije koje su razmijenjene.

(IEEE 90)

Ove dvije definicije opisuju razinu proizvoda informacijske i komunikacijske tehnologije i usluga koja se korisniku pruža kad se njima koristi. Pri tome, druga definicija izričito govori o razmjeni i ponovnom korištenju informacija od strane pojedinih sustava.

... da bi se bilo interoperabilnim,

treba se aktivno baviti stalnim procesom osiguravanja takva načina upravljanja sustavima, procedurama i kulturom neke ustanove da se povećaju mogućnosti razmjene i ponovne upotrebe informacija unutar te ustanove ili među takvim ustanovama.

(Paul Miller, *Interoperability: What is it and Why should I Want it?*)

Stvarno je interoperabilna ona ustanova koja može povećati vrijednost i ponovnu upotrebu potencijala informacija kojima upravlja. Takva ustanova može i efikasno razmjenjivati te informacije s drugim, isto tako interoperabilnim, tijelima, dopuštajući da se pri tom stvaraju nova znanja iz identifikacije odnosa između skupova podataka koji do tada nisu bili povezani ni na koji način.

(Paul Miller, *Interoperability: What is it and Why should I Want it?*)

Obje definicije, pri čemu je druga konkretizirana definicija prve načelne, stavlja u središte koncept (smjer) razvoja neke ustanove - upravljanje ustanovom i njenim "dobrima", njezinu funkcionalnost. Ključne su riječi ovdje "razmjena" i "ponovna upotreba" te "potencijal informacija" i "stvaranje novih znanja".

Interoperabilnost

jest gradnja suglasnosti i standarda i praktičnih projekata i usluga. Pri tome je interoperabilnost sadržaja najteži dio. Praktični je izazov: baviti se heterogenošću.

(Traugott Koch, *Practical Experiences with Interoperability and Heterogeneity in Digital Library Projects*)

Gornja je definicija najrelevantnija za temu ovoga rada i općenito seminara. Ona upućuje, s jedne strane, na mogući prostor zajedničke izgradnje "sadržaja" – metodologiju stvaranja (meta)podataka o građi a, s druge strane, na mogući prostor zajedničke izgradnje pristupa tom "sadržaju" – metodologiju stvaranja putova/pristupa do (meta)podataka o građi. Ipak, bavljenje tim konceptom potrebno je staviti u kontekst dviju prethodno navedenih definicija: informacijske i komunikacijske tehnologije i djelovanja ustanova kao otvorenih sustava.

Interoperabilnost će se, u posljednje navedenom smislu, promatrati na četiri razine toka i upravljanja informacijama. To su razina protokola između računalnih sustava, struktura zapisa, označitelji sadržaja zapisa i sadržaj (meta)podataka zapisa.

Tri posljednje navedene razine spadaju u područje strojno čitljivih formata zapisa za razmjenu podataka i one su temeljito analizirane i dobro poznate u dijelu knjižničarstva koji se bavi strojno čitljivom katalogizacijom i MARC formatima.¹ U okviru ovih seminara već se raspravljalo o komunikaciji – konverziji (prevodenju) formata zapisa koji se koriste u trima djelatnostima, kao i njihove konverzije u odnosu na internetski format Dublin Core. Govorilo se i o razini protokola za komunikaciju između različitih sustava. U ovom će se radu opis sve četiri razine proširiti stavljanjem u kontekst novorazvijenih standarda za obilježavanje i razmjenu sadržaja (meta)podataka u internetskom okruženju. Posebna će se pažnja posvetiti referiranju usvojene, dobro poznate terminologije na terminologiju koja se razvila s novim standardima. Ta se nova terminologija u prvom redu služi pojmovima i konceptima iz lingvistike.² Ona je stručnjacima iz najrazličitijih područja koji se bave formalizacijom intelektualne organizacije informacija na internetu, pružila dovoljno apstraktan (tj. nekontaminiran pojedinom strukom) alat za precizno imenovanje pojava, njihovo uređivanje i opis njihove interakcije.

Razina 1: protokoli između računalnih sustava

O međunarodnom standardu za komunikaciju između računalnih sustava u više se navrata i na različite načine govorilo na seminarima *Arhivi, knjižnice, muzeji*. Stoga se ovdje neće ulaziti u njegov detaljan opis, već se upućuje na članke Paula Millera Z39.50 za sve³ i Tinke Katić Z39.50 u sustavu OLIS (*Oxford Library Information System*).⁴

¹ Vidjeti, npr., Willer, Mirna. UNIMARC u teoriji i praksi. Rijeka : Benja, 1996.

² Vidjeti, npr., Svenonius, Elaine. The intellectual foundation of information organization. Cambridge, Mass. : The MIT Press, 2000.

³ Miller, Paul. Z39.50 za sve. // Vjesnik bibliotekara Hrvatske 43, 1-2(2000), 69-83.

⁴ Katić, Tinka. Z39.50 u sustavu OLIS (*Oxford Library Information System*). // 2. i 3. seminar Arhivi, knjižnice, muzeji : mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova / uredile Mirna Willer i Tinka Katić. Zagreb : Hrvatsko muzejsko društvo, 2000. Str. 139-148.

Američki standard ANSI/NISO Z39.50-1995 Information Retrieval⁵ prihvaćen je kao međunarodni standard ISO 23950:1998 Information and Documentation – Information Retrieval,⁶ a Hrvatski zavod za normizaciju i mjeriteljstvo pokrenuo je postupak prihvatanje toga standarda kao hrvatske norme.

Taj je standard, ukratko, mrežni protokol (protokol = strogi skup pravila koji upravlja razmjenom informacija između računala) koji omogućuje pretraživanje (uglavnom udaljenih) heterogenih baza podataka i pronalaženje podataka putem jednokorisničkog sučelja. Najčešće se koristi za pronalaženje bibliografskih podataka (komunikaciju između računalnih sustava kao što su oni koji se rabe za upravljanje knjižničnim katalozima), ali i za ne-knjižnične primjene. Takvi se sustavi obično zovu "otvorenim sustavima" jer, s jedne strane, stavljaju svoje informacije na raspolaganje velikom broju korisnika preko npr. OPAC-a, a, s druge strane, spremni su komunicirati, tj. razmjenjivati te informacije s drugim, isto tako otvorenim, sustavima. U toj komunikaciji računalo (korisnik: **izvor**) šalje zahtjev (**pretraživanje**) drugom računalu (poslužitelju: **cilju**) koji tada opslužuje zahtjev (**pronalaženje**) i vraća neku vrstu odgovora (prikaz rezultata: **predstavljajanje**). Već se na ovoj razini razmatranja interoperabilnosti uočava "pomicanje" terminologije: korisnik (*client*) se naziva izvorom (*origin*), a poslužitelj (*server*) ciljem (*target*); pretraživanje (*search*) i pronalaženje (*retrieval*) upotrebljavaju se za iste postupke korištene u prethodnoj tehnologiji pretraživanja strojno čitljivih kataloga (OPAC-a jednog sustava), ali se specificira i korak (funkcija) prikaza rezultata dobivenih od različitih sustava – predstavljanje (*display*). Predstavljanje zapisa može biti u nekoj od struktura = **sintaksa** zapisa (o "sintaksi" kasnije), kao npr. UKMARC umjesto UNIMARC-a ili Dublin Core ili SUTRS (o "SUTRS" kasnije).

Kao i svaki standard, tako se i ISO 23950 odnosno Z39.50 može primijeniti na različite načine definirane profilima. Profil je:

*posebna interpretacija načina na koji se treba koristiti Z39.50 ili njegov podskup kako bi se udovoljilo potrebama određene aplikacije (GILS: The Government Information Locator Service - državne informacije), zadaće (ATS-1 za pretraživanje po autoru, naslovu i predmetu), zajednice (CIMI profil) ili okruženja (CENL profil za europske nacionalne knjižnice)).*⁷

U okviru knjižnične zajednice definiran je profil Bib-1 za pretraživanje bibliografskih podataka. Taj je profil krajem 1990-ih razrađen u takozvani Bath profil, prema gradu Bathu u Velikoj Britaniji u kojem su se okupili stručnjaci i nastavili rad na njegovoj standardizaciji vodeći pri tom računa da se profil proširi i na druge djelatnosti koje se bave pretraživanjem i pronalaženjem podataka (*resource discovery*), a ne samo na knjižnice.

⁵ Temeljni članak za razumijevanje Z39.50 svakako je onaj Clifforda A. Lyncha The Z39.50 information retrieval standard : Part 1 : A strategic view of its past, present and future. // D-lib magazine (April 1997) <http://www.dlib.org/dlib/april97/04lynch.html> (1998-05-18).

⁶ Međunarodni standard ISO 23950 zamijenio je međunarodni standard ISO/OSI 10163-1, Međunarodni standard za povezivanje otvorenih sustava: ISO 10163-1:1993 Information and Documentation - Open Systems Interconnection - Search and Retrieve Application Protocol Specification - Part 1.

⁷ Miller, Paul. Z39.50 za sve. // Vjesnik bibliotekara Hrvatske 43, 1-2(2000), 80.

Izvorni je naziv profila The Bath Profile: An International Z39.50 Specification for Library Applications and Resource Discovery, a verzija 1.1 objavljena je u lipnju 2000.

Bath profil⁸ podržava tri funkcionalne skupine tj. vrste postupaka (pretraživanje – pronalaženje) koji manipuliraju određenim vrstama podataka u procesu komunikacije između sustava:

Funkcionalna skupina A za osnovno pretraživanje i pronalaženje bibliografskih podataka, s osnovnim naglaskom na knjižnične kataloge: osnovni mehanizmi pretraživanja i pronalaženja koje treba knjižnični korisnik kad koristi knjižnični katalog ili druge elektroničke izvore. Definira dodatna pretraživanja za preciznija pretraživanja kako bi se podržale druge korisnikove potrebe u pretraživanju.

Funkcionalna skupina B za pretraživanje i pronalaženje podataka o fondu: definicija zahtjeva za pretraživanje i pronalaženje za bibliografske informacije i informacije o fondu dovoljne za identifikaciju knjižničnog fonda.

Funkcionalna skupina C za pretraživanje i pronalaženje među različitim domenama: definicija mehanizama za pretraživanje i pronalaženje kako bi se obuhvatilo pretraživanje informacija među područjima djelatnosti tj. domenama.

Profil predviđa mogućnost definiranja drugih funkcionalnih skupina u svojim daljnjim verzijama, kao što su funkcionalna skupina za ažuriranje skupnog kataloga i funkcionalna skupina za naručivanje bibliografske jedinice i dostavu dokumenta.

Namjena je Bath profila, kako je opisana dokumentom, prepoznati ona svojstva standarda Z39.50 koja su potrebna da se omogući efikasna upotreba Z39.50 softvera u različitim knjižničnim primjenama, uključujući pretraživanje i pronalaženje bibliografskih podataka iz knjižničnih kataloga; prijenos informacija o fondu; pretraživanje “preko” domena između knjižnica, muzeja i arhiva; ažuriranje skupnih kataloga; naručivanje bibliografske jedinice i dostava dokumenta. Naime, u opisu se profila kaže da pronalaženje informacija nije samo sebi svrhom, već je to prije svega jedan korak prema široj djelatnosti kao što je međuknjižnična posudba, katalogizacija, referentna djelatnost, nabava ili ažuriranje kataloga (skupni katalog).

Korisnost je profila dvojaka. Prvo, namjena mu je da definira osnovni skup funkcionalnosti i Z39.50 specifikacija kako bi se omogućilo *međunarodno* ili *izvanmacionalno* pretraživanje i pronalaženje, posebno kad Z-klijent nema detaljne informacije o jednom ili više Z-poslužitelja. Drugo, specifikacije profila daju temelje za interoperabilnost između Z-klijenata i Z-poslužitelja izvan primarne jurisdikcije regionalnih, nacionalnih, državnih, lokalnih profila ili profila definiranih za pojedine projekte kad njihovi pojedinačni profili uključuju specifikacije Bath profila. Namjena je osnovne funkcionalnosti i specifikacija definiranih ovim profilom da posluže kao stvarni podskup navedenih profila.

⁸ Opis profila koji slijedi preuzet iz: The Bath profile : an international Z39.50 specification for library applications and resource discovery. Release 1.1. Released June 2000. Internationally registered profile / developed by The Bath Group; edited by The Bath Profile Editorial Team Carrol Lunau, Paul Miller, and William Moen. Released June 2000. <http://www.ukoln.ac.uk/interop-focus/bath/current> (2000-11-07).

Bath profil definira:

- osnovno bibliografsko pretraživanje: autor, naslov, predmet, ključna riječ, Booleova pretraživanja i kraćenje,
- osnovno bibliografsko pronalaženje: bibliografski zapisi i zapisi o fondu iz knjižničnih kataloga u skraćenom ili potpunom MARC formatu, ali i slobodnom tekstu kako bi se osigurala interoperabilnost (MARC21, UNIMARC i slobodan tekst), i
- pretraživanje i pronalaženje između domena (knjižnični katalogi, vladine informacije, muzejski sustavi i arhivi): postiže se nestrukturiranim slobodnim tekstom – SUTRS: Simple Unstructured Text Record Syntax i XML-om; profil podrazumijeva da je korisnikovo ponašanje pri pretraživanju različitih domena slično ponašanju pri pretraživanju definiranom za osnovno bibliografsko pretraživanje.

Profil definira i razine suglasnosti (*conformance level*) za svaku pojedinu funkcionalnu skupinu, a to je specifikacija i ponašanje Z39.50 korisnika i poslužitelja. Za sve tri funkcionalne skupine definirane su tri razine: 0, 1 i 2.

U Funkcionalnoj skupini A suglasnost s Razinom 0 omogućuje ono što se u nekim WebPAC-ima naziva jednostavnim pretraživanjem: to je osnovna razina pretraživanja po autoru, naslovu, predmetnici i ključnoj riječi. Naslov, predmet i opća ključna riječ pretražuju se kao ključne riječi da bi se osigurao najveći mogući odaziv, dok se autor pretražuje kao fraza, po mogućnosti normalizirana, kako bi se postigla visoka preciznost pretraživanja.⁹ U pronalaženju (dakle, odgovoru cilja) Z-korisnik treba podržavati MARC format UNIMARC ili MARC21, SUTRS i XML, dok Z-poslužitelj treba podržavati UNIMARC ili MARC21 i SUTRS ili XML. U istoj skupini, suglasnost s Razinom 1 omogućuje, uz osnovno, i preciznije pretraživanje. Osim pretraživanja po autoru, naslovu i predmetnici, može se pretraživati po standardnom broju i datumu izdavanja, a pretraživanje po ključnoj riječi pospješeno je kraćenjem riječi. U odnosu na Razinu 0, na ovoj razini Z-klijent treba podržavati UNIMARC i MARC21, uz SUTRS i XML.

Razina 2 nije specificirana ovim profilom, ali se navodi popis preliminarnih vrsta pretraživanja po različitim parametrima, kao npr. ISBN, ISSN, Lokalni broj, DOI/URN, jezik, klasifikacijska shema itd. Također nisu specificirane razine za Funkcionalnu skupinu B (bibliografski podaci i podaci o fondu).

Razine 0 i 1 su, međutim, specificirane za Funkcionalnu skupinu C koja je nama posebno zanimljiva jer specificira komunikaciju između domena uključujući knjižnične kataloge, vladine informacije, muzejske sustave i arhive. Da bi se omogućilo pronalaženje u ovakvom kontekstu, profil zahtijeva SUTRS i XML za osnovnu razmjenu podataka, pri čemu će se sintaksa XML zapisa koristiti s DTD-om za 15 elemenata jednostavnog Dublin Corea.¹⁰ DTD za Dublin Core opisan je u Dodatku D Bath profila: *eXtensible Markup Language (XML) Document Type Definition (DTD) for Dublin Core Simple*.

⁹ Osnova za ovakvu specifikaciju pretraživanja po autoru iskustva su u radu s ranijim profilima, posebno ATS-1. Vidjeti i rezultate rada na profilu projekta ONE koji je opisala Liv Aasa Holm u članku Normativna kontrola u međunarodnom kontekstu u novome okruženju. // Vjesnik bibliotekara Hrvatske 43, 1-2(2000), 23-29.

¹⁰ Jednostavan je Dublin Core (Dublin Core Simple) DC bez upotrebe kvalifikatora. Za pregledan prikaz DC-a i ulogu kvalifikatora vidjeti Hakala, Juha. Dublinski osnovni skup elemenata metapodataka = Dublin Core Metadata Element Set. // Vjesnik bibliotekara Hrvatske 43, 1-2(2000), 49-68 i Willer, Mirna. Sedma radionica o Dublinskom osnovnom skupu elemenata metapodataka = The 7th Dublin Core Metadata Workshop, Frankfurt am Main, 25.-27. listopada 1999. // Vjesnik bibliotekara Hrvatske 43, 1-2(2000), 135-140.

Razine su definirane za ovu funkcionalnu skupinu kao i za Funkcionalnu skupinu A, samo što nazivi i upotreba pretraživanja odražavaju općenitiji pristup (od onoga u Funkcionalnoj skupini A koji je obilježen knjižničarskim pristupom) i semantiku koja se odnosi na skup elemenata podataka Dublin Corea. Unutar Razine 0 pretražuje se po stvaratelju (DC.Creator), naslovu (DC.Title), temi (DC.Subject) i bilo kojem polju (*any search*), i to kao ključna riječ (cijela riječ), dok je unutar Razine 1 omogućeno detaljnije pretraživanje. Navedenim vrstama dodani su standardni identifikator i datum izdavanja (kao i u Funkcionalnoj skupini A), a pretražuje se po ključnoj riječi i frazama koje se koriste kao pristupnice (npr. naslov ili pristupnica kako je definira određeni poslužitelj), a omogućeno je i kraćenje.¹¹

Zaključno: želimo li međusobno komunicirati ili, točnije, želimo li da naša računala pretražuju različite izvore informacija, moramo se u svakom slučaju pripremiti na dvije vrste komunikacije. Da bi međusobno komunicirale, knjižnice (u Hrvatskoj, ali i prema svijetu) u prvom koraku trebaju udovoljiti specifikacijama u Funkcionalnoj skupini A, Razini 0, a sljedeći je korak Razina 1. Da bi komunicirale s muzejima i arhivima, trebaju za početak udovoljiti specifikacijama Funkcionalne skupine C i Razini 0. Muzejima i arhivima koji ne koriste MARC formate "preostaje samo" komunikacija na Razini 0 u obje funkcionalne skupine (A i C) pomoću SUTRS-a i XML-a i/ili Dublin Corea i XML-a. Je li to zadovoljavajuće za muzeje i arhive ako se zna da su i Dublin Core i SUTRS ograničeni skup elemenata podataka (podskup) jednog semantički mnogo bogatijeg načina opisivanja građe/gradiva? Kojim će jezikom oni komunicirati unutar "svog" informacijskog prostora?

Razina 2: struktura zapisa = sintaksa

Struktura strojno čitljivoga zapisa fizičko je (tvarno) predstavljanje i prikaz podataka. Termin struktura zapisa poistovjećuje se s terminom sintaksa. Ova je razina definirana sljedećim standardima:

ISO 2709:1981 *Format za razmjenu bibliografskih podataka na magnetskoj vrpici* osnova je MARC (MACHine Readable Cataloguing) formatima kao što su UNIMARC, MARC21, UKMARC itd. Standard tako propisuje da svaki format tj. svaka primjena (određena specifikacija) toga standarda treba imati oznaku zapisa, kazalo, određenu dužinu koda (označitelja sadržaja) za opis polja podataka, dva indikatora, dva znaka za obilježavanje oznake potpolja itd., te da se bibliografski podaci (elementi podataka) zajedno s tim označiteljima sadržaja trebaju strukturirati tj. "poredati" na magnetsku vrpcu, na određeni način.

ISO 8879:1986 *Standard Generalized Markup Language – SGML* standardan je opći jezik za obilježavanje elektroničkog teksta, a razvila ga je zajednica nakladnika i izdavača. Što znači obilježavanje? Povijesno, ta se riječ koristila za opisivanje anotacija ili drugih oznaka unutar teksta koje su upućivale slagara kako otisnuti ili oblikovati određeni odjeljak ili naslov. Na primjer, valovita crta znači masna slova, simbol √ umetanje teksta itd. Auto-

¹¹ Vidjeti u Dodatku izvornu tablicu koja sažima tipove pretraživanja za Funkcionalne skupine A i C te Razine 0 i 1.

matizacijom oblikovanja i tiskanja značenje termina prošireno je na različite vrste kodova za obilježavanje koji su se uvrštavali u elektroničke tekstove kako bi upravljali oblikovanjem, tiskanjem ili drugim vrstama obrade. Prema tome, obilježavanje se definira kao neki od načina pojašnjavanja interpretacije teksta. Jezik za obilježavanje mora specificirati koje je obilježavanje dopušteno, odnosno, kako se obilježavanje razlikuje od teksta i što znači obilježavanje.¹² Prve tri specifikacije određuje SGML, dok četvrtu određuju korisnici opisujući određenu primjenu toga standarda.

Ta je primjena definirana profilom, interpretacijom načina upotrebe SGML-a za određenu svrhu, takozvanim DTD-ima, objašnjenjem značenja upotrebljenih kodova pri interpretaciji teksta. DTD je *Document Type Definition* (definicija tipa dokumenta). Dakle, možemo prepoznati terminološko izjednačenje pojmova:

DTD = format

tj. definirana specifikacija strukture i obilježavanja/označivanja.

Navedimo nekoliko DTD-a o kojima se na seminarima već govorilo:

- EAD.DTD – Encoded Archival Description: DTD koji koristi strukturu SGML-a; definiran je za arhivističku zajednicu; definira označitelje za obavijesna pomagala (*finding aid* = kataložni zapis), a ne same dokumente.
- TEI .DTD – Text Encoding Initiative: DTD izrađen za vrlo složena obilježavanja tekstualnih dokumenata, a podržava i tekstualnu analizu dokumenata. Korišten je u mnogim projektima digitalizacije, kao npr. *Making of America* i *American Memory* Kongresne knjižnice.
- HTML – HyperText Markup Language: DTD koji koristi strukturu SGML-a; standardizirani skup polja koji se najviše koristi u web-okruženju: web-pretraživači ga «lako» prikazuju jer je izrađen upravo s namjerom da omogućiti prezentacijske aspekte dokumenta.
- XML - eXtensible Markup Language: podstruktura SGML-a za web-dokumente.

XML je relativno nova sintaksa jezika za obilježavanje elektroničkih tekstova te je potrebno ovdje opisati njegova najosnovnija svojstva:

- XML ne zahtijeva formalni DTD, već razmjerno umanjenu “shemu” tj. interpretaciji vrijednosti nekog označitelja sadržaja (o “označitelju sadržaja” odmah dalje): XML oznake nemaju unaprijed određeno značenje;¹³
- obilježen dokument s definiranim skupom polja koji se koristi u XML-u mora, zbog weba, pratiti takozvani *style sheet* koji će pretraživaču definirati njegov prikaz;
- *style sheet* omogućuje da se ne obilježava prikaz kao u HTML-u, već da se identificiraju elementi podataka;
- XML ne definira skup polja koja se moraju primjenjivati (semantiku: o “semantici” odmah dalje), već strukturu podataka koja je otvorena za definiranje bilo kojeg skupa oznaka polja (*tag, content designator*) koji definira određeni skup elemenata podataka, kao na primjer Dublin Core ili MARC.

¹² Parafrazirano iz poglavlja A gentle introduction to SGML. // TEI guidelines for electronic text encoding and interchange. <http://etext.virginia.edu/bin/tei-tocs?div=DIV1&id=SG> (2001-01-26).

¹³ Za razliku od HTML-a koji je sam DTD.

Zbog svojstava da definira samo strukturu, za XML može vrijediti sljedeće terminološko izjednačenje pojmova:

XML = ISO 2709

Kako bi omogućila interoperabilnost formata na razini sintakse, Kongresna je knjižnica izradila SGML DTD za MARC21 zapise i konverzijske programe između struktura MARC21/ISO 2709 i SGML/ISO 8879. Sada se radi na specifikaciji XML DTD-a za MARC21: XML oznaka polja jednaka je oznakama polja MARC-a, tj. upotrebljavaju se ista polja kao i u strukturi formata MARC21, npr.:

Primjer 1 – MARC/2709

<245> MARC21 (dio unosa u kazalo)
10\$aData on the web :\$bfrom relations to semistructures data and XML /\$cSerge Abiteboul

Primjer 2 – MARC/XML

<mrcb245 i1=»1» i2=»0»/> <mrcb245-a>Data on the web :</ mrcb245-a>
<mrcb245-b>from relations to semistructures data and XML </ mrcb245-b>
<mrcb245-c>Serge Abiteboul</ mrcb245-c>

(mrcb = MARC bibliographic)

(i = indikator)¹⁴

Ti bi primjeri u UNIMARC-u izgledali ovako:

Primjer 1 – MARC/2709

<200> UNIMARC (dio unosa u kazalo)
#1\$aData on the web\$efrom relations to semistructures data and XML\$fSerge Abiteboul

Primjer 2 – MARC/XML

<unmb200 i1=»#» i2=»1»/> <unmb200-a>Data on the web</ unmb200-a>
<unmb200-e>from relations to semistructures data and XML</ unmb200-e>
<unmb200-f>Serge Abiteboul</ unmb200-f>

(unmb¹⁵ = UNIMARC bibliographic)

¹⁴ Primjeri preuzeti iz: MacCallum, Sally. Extending MARC for bibliographic control in the web environment : challenges and alternatives. // Bicentennial Conference on Bibliographic Control for the New Millennium Confronting the challenges on networked resources and the web, November 15-17, 2000 / sponsored by the Library of Congress Directorate. http://lcweb.loc.gov/catdir/bibcontrol/mccallum_paper.html (2000-10-04) *Napomena:* Interpunkcija na kraju svakog elementa podatka, npr. dvotočka nakon glavnoga stvarnog naslova i kosa crta nakon podatka o odgovornosti, način je zadržavanja interpunkcije ISBD-a uz bibliografske podatke karakterističan za format MARC21. Dakle, ta interpunkcija nema nikakve veze s interpunkcijom koju koristi ili zahtijeva XML. Primjer u UNIMARC-u ne sadrži interpunkciju.

¹⁵ Kod "unmb" je ilustrativan.

Francusko Ministarstvo kulture i komunikacija, Služba za istraživanje i tehnologiju (Ministère de la culture et de la communication, Mission de la recherche et de la technologie) izradilo je projekt XML DTD za UNIMARC formate pod nazivom BiblioML (UNIMARC/Bibliographic Format) i AuthoritiesML (UNIMARC/Authorities Format) te konverzijske programe između struktura UNIMARC/ISO 2709 i UNIMARC/XML. IFLA-in Stalni odbor za UNIMARC (Permanent UNIMARC Committee) razmatrat će taj dokument i eventualno ga proglasiti preporukom za upotrebu UNIMARC-a u sintaksi XML-a.

Posebno valja naglasiti svojstvo XML-a da se može koristiti i kao struktura za unos i pohranu, a ne samo za razmjenu podataka, što nije slučaj s ISO 2709 koji je isključivo definiran kao format za razmjenu. Također, alati MARC-u-XML-u omogućit će da dokumenti katalogizirani u MARC-u budu konvertirani iz MARC sustava (bibliografske baze podataka) u MARC/XML te budu uključeni u XML elektronički dokument, kao i obrnuto: da elektronički dokument sa svojim opisom u MARC/XML-u bude učitani i pohranjen u bazu bibliografskih podataka.¹⁶ Ograničenja na razini sintakse MARC sustava tj. strukture ISO 2709, koja se naročito odnose na dužinu zapisa i bilježenje relativnih odnosa među zapisima te odnos zapisa prema elektroničkom tekstu koji opisuju, time su ukinuta čime je omogućena evolucija MARC-a kao formata, a ne revolucija – prijelaz na neki drugi format zapisa (kojeg se knjižničari plaše).

Kada se govori o sintaksi, potrebno je spomenuti još dva internetska standarda. To su RDF i SUTRS.

RDF: *Resource Description Framework* (Okvir za opis elektroničke građe) razvio je World Wide Web Consortium (W3C). To je opći sintaksički model tj. infrastruktura koja omogućuje sadržajno obilježavanje, razmjenu i ponovnu upotrebu strukturiranih metapodataka koji mogu biti modularno definirani, te njihovih odnosa - temeljnog građevnog bloka «semantičke mreže». RDF ne propisuje semantiku, već omogućuje pojedinoj zajednici koja opisuje podatke da sama definira semantičku strukturu koja najbolje odražava zahtjeve te zajednice. XML se pri tome koristi kao opća sintaksa za razmjenu i obradu metapodataka.

RDF *izjava* definira se kao

“određen izvor zajedno s imenovanim svojstvom plus vrijednost toga svojstva za taj izvor. Navedena tri odvojena dijela izjave zovu se *subjekt*, *predikat* i *objekt*. Objekt izjave (tj. vrijednost svojstva) može biti neki drugi izvor ili može biti doslovno svojstvo tj. izvor (specificiran URI-jem¹⁷) ili jednostavan niz ili neki drugi primitivan tip podatka definiran XML-om. U terminologiji RDF-a, *doslovno svojstvo* može imati sadržaj koji je obilježen XML-om, ali ga RDF procesor dalje ne vrednuje.”¹⁸

U pojašnjenju uloge RDF-a, može pomoći i opis kompatibilnosti RDF-a i Dublin Corea koju objašnjava Thomas Baker. Dakle,

¹⁶ Izričito se preporuča članak Dicka R. Millera o ulozi XML-a u daljnjem razvoju bibliotečnih sustava i formata: Miller, Dick R. XML : libraries' strategic opportunity. // Library journal (2000) <http://xmlmarc.stanford.edu/LJ/> (2000-10-18).

¹⁷ URI: Universal Resource Identifier.

¹⁸ Resource Description Framework (RDF) model and syntax specification / edited by Ola Lassila and Ralph Swick. <http://www.w3.org/TR/REC-rdf-syntax> (2001-01-27).

RDF pruža opći model za iskaze, dok Dublin Core pruža određen tip dijalektalne izjave o informacijskim izvorima pri čemu dopušta upotrebu malog skupa posebnih riječi. Za izražavanje svoga konceptualnog modela, RDF pruža specifično kodiranje u XML-u, za razliku od Dublin Corea, koji je po svom dizajnu nezavisan od bilo kojeg određenog kodirajućeg formata. RDF je, dakle, samo jedan od mogućih informacijskih modela koji mogu koristiti oznake iz Dublin Corea, dok je Dublin Core samo jedan od jezika koji se mogu izraziti u RDF-u.¹⁹

SUTRS: *Simple Unstructured Text Record Syntax* jednostavna je nestrukturirana sintaksa tekstualnih zapisa koja se koristi u slučajevima razmjene podataka kada, zbog pomanjkanja upotrebe strukturiranih podataka u izvornom ili ciljnom sustavu ili zbog heterogenosti sustava koji komuniciraju (npr. arhivi, knjižnice, muzeji), nije moguće "razgovarati" jednom od opisanih sintaksi. (Vidjeti ponovno gore Bath profil, Funkcionalna skupina C.)

Razina 3: označitelji sadržaja zapisa = semantika

Označitelji sadržaja zapisa čine sustav kodova koji identificiraju elemente podatka u zapisu i pružaju o njima dodatne informacije. Na ovoj razini možemo govoriti o sljedećem terminološkom izjednačenju pojmova:

obilježavanje (*markup*) = skup polja (*field; tag*) = označitelj sadržaja (*content designator*)

tj. sustav simbola kojim se na određeni način identificiraju podaci

zapis (*record*)

tj. skupina informacija koja opisuje građu

Standardi označitelja sadržaja zapisa (standardizirana semantika) definirani za upotrebu u određenoj specijaliziranoj djelatnosti jesu sljedeći:

- MARC s «profilima», npr. UNIMARC, MARC21, UKMARC... za određene vrste podataka: Authorities/Bibliographic/Holdings/Classification Data,
- Dublin Core s profilima, kao npr. CIMI:DCMES, BIBLINK DC, Renardus DC itd.

Kao što je već spomenuto, konverzija ili «prevođenje» između semantika ili označitelja sadržaja tih formata već je vrlo ustaljena. Uobičajena prevođenja su na primjer:

- MARC21 – UKMARC (i obrnuto),
- MARC21 – UNIMARC (i obrnuto),
- Dublin Core – MARC21 (i obrnuto),
- Dublin Core – UNIMARC (i obrnuto) itd.

Kao primjer može poslužiti usluga preuzimanja kataložnog zapisa (*downloading*) iz CERL-ove (Consortium of European Research Libraries) baze podataka Hand Press

¹⁹ Baker, Thomas. A grammar for Dublin Core. // D-lib magazine 6, 10(October 2000). <http://www.dlib/october00/baker/10baker.html> (2000-10-18). Vidjeti dalje objašnjenje odnosa između RDF-a i XML-a u: Berners-Lee, Tim. Why RDF model is different from the XML model. Last modified 1998-10-14. <http://www.w3.org/DesignIssues/RDF-XML.html> (2000-11-07).

Book u lokalnu bazu, pri čemu se koristi posebno razvijen softver ICON. ICON - ZedMARC box omogućuje konverziju formata MARC21 u UNIMARC u okruženju Z39.50 odnosno Bath profila.²⁰

Razina 4: sadržaj podatka (podataka) zapisa = (kataložni) sadržaj = metapodatak (metapodaci)

Na ovoj, "najnižoj", razini govorimo o samom sadržaju metapodataka zapisa koji je obično standardiziran unutar pojedine djelatnosti nacionalnim standardima koji se uglavnom temelje na međunarodnima. Međutim, sadržaj metapodataka zapisa najčešće je posljedica kulturnih i jezičnih osobitosti a time i razlika, nacionalnih i lokalnih tradicija bilo u proizvodnji «jedinice građe» bilo u njenom opisivanju, povijesnih prethodnika u stvaranju informacijskih pomagala, tehnološkom stupnju razvoja te potreba i navika korisnika o kojem brine određena ustanova.

Dakle, ovdje govorimo o najmanjoj jedinici informacije i njezinoj vrijednosti:

element podatka

tj. pojedinačan, najmanji dio bibliografskog podatka koji se može zasebno identificirati, uključujući elemente koji se obično nazivaju atributima ili kvalifikatorima

atribut ili kvalifikator = vrijednost

Atribut modificira svojstva elemenata podataka specificiranjem, na način kako to čine pridjevi u prirodnom jeziku, "koja je vrsta" predmeta, datuma ili odnosa. Jedna je vrsta kvalifikatora takozvana shema za kodiranje (*encoding scheme*), tj. pokazivač kontekstualne informacije ili pravilo raščlanjivanja (kataložna pravila, shema predmetnog sustava, ISO standard za način bilježenja određene vrste podatka, npr. datum) koji pomaže u interpretaciji vrijednosti elementa.²¹

Na primjer, ako u potpolju UNIMARC-a 210\$d imamo sadržaj "1879. [i.e. 1897.]", a u potpolju 801\$g imamo sadržaj "PPIAK", to znači da je element podatka Datum izdavanja (210) zabilježen prema kataložnim pravilima *Pravilnika i priručnika za izradbu abecednih kataloga* Eve Verone, odnosno: na predlošku je zabilježena godina 1879., ali je ispravna godina izdavanja 1897. što je ustanovljeno prema nekom drugom izvoru podataka, a ne prema impresumu. Ili, ako u polju Dublin Core dc:date tj. Datum, imamo sadržaj "2001-01-06", a u dcq tj. *Dublin Core Qualifier*, imamo sadržaj "iso8601", to znači da je element podatka Datum zabilježen prema međunarodnom standardu ISO 8601 i znači: 6. siječnja 2001. Da nema kvalifikatora, moglo bi se nagađati da je vrijednost elementa Datum 1. lipnja 2001.

²⁰ Više o tom softveru vidjeti na: Consortium of European Research Libraries <http://www.cerl.crxnet.com> i na Crossnet System Limited <http://www.crxnet.com>.

²¹ Definicija je djelomično preuzeta, a djelomično parafrazirana iz: Baker, Thomas. A grammar for Dublin Core. // D-lib magazine 6, 10(October 2000). <http://www.dlib/october00/baker/10baker.html> (2000-10-18).

Zaključak

Optimalna definicija interoperabilnosti sadržaja mogla bi biti: omogućiti korisniku da dobije željenu informaciju o objektu na jedinstven način²² bez obzira gdje se ona nalazi (koji sustav daje informaciju), i na kojem je mediju pohranjena. Koje probleme uočavamo pri realizaciji tako opisane interoperabilnosti? Nabrojimo samo neke:

- što treba biti interoperabilno, a što ne,²³
- kako definiramo entitete koje opisujemo,
- koji su nam zajednički elementi (meta)podataka,
- kako definiramo obvezne, izborne, ponovljive, neponovljive elemente podataka,
- kako rješavamo granularnost opisa,
- kako opisujemo verzije jedinice i odnose među njima,
- koje kriterije kvalitete tj. vrijednosti sadržaja (pravila, standarde, popise, sheme itd.) možemo dijeliti,
- kako održavati interoperabilnost uz promjenu i doradu sintakse i semantike zapisa,
- kako održavati indeksne strukture informacijskog sustava da bi se omogućilo efikasno i potpuno pretraživanje protokolom Z39.50 (npr. za probleme koji nastaju pri pretraživanju po autoru vidjeti spomenuti članak Liv Aase Holm, a za pretraživanje u sustavima koji nemaju ujednačene postupke katalogiziranja i indeksiranja vidjeti preporuke radne grupe škotskog *gatewaya* CAIRNS²⁴),
- koji je obuhvat (meta)podataka u odnosu na specifične/potencijalne korisnike itd.

Ipak, realno je pitanje: je li moguće osigurati takvu interoperabilnost sadržaja metapodataka?

²² Ovdje se govori samo o metapodacima: pretpostavka je da je krajnji cilj svakog sustava osigurati dostup objektu, a u kontekstu digitalno rođene elektroničke građe ili digitalizirane građe u internetskom okruženju osigurati uvid u sam objekt.

²³ Vidjeti predavanje *Profiling ALM applications* Liv Aase Holm održano na 24. seminaru Library Systems Seminar / European Libraries Automation Group, Paris, 2000.

²⁴ CAIRNS Project recommendations for a cataloguing and indexing strategy for Scottish libraries. Glasgow : CAIRNS Cataloguing and Indexing Working Group, 2000. <http://cairns.lib.gla.ac.uk/docs/CAIRNSCatStrat.pdf>. CAIRNS je kratica za Co-operative Academic Information Retrieval Network for Scotland.

DODATAK

The following table summarises the search types defined for Functional Areas A and C and Levels 0 and 1

Fields Containing One or More Words				
Search Type	Functional Area A		Functional Area C	
	Level 0	Level 1	Level 0	Level 1
Keyword	Title, Subject, Any	Author, Title, Subject, Any	Author, Title, Subject, Any	Author, Title, Subject, Any
Keyword with right truncation		Author, Title, Subject, Any		Author, Title, Subject, Any
First characters		Title, Subject		
First words		Title, Subject		
Unanchored normalized name	Author	Author		
Unanchored phrase				Author, Title, Subject, Any
Exact match		Author, Title, Subject		

One Word Fields				
Search Type	Functional Area A		Functional Area C	
	Level 0	Level 1	Level 0	Level 1
Exact match		Identifier; Date of Publication		Identifier; Date of Publication

The Holdings Attribute Set is used for searches involving holdings information and is detailed in Functional Area B.

Literatura

- Baker, Thomas. A grammar for Dublin Core. // D-lib magazine 6, 10(October 2000). <http://www.dlib/october00/baker/10baker.html> (2000-10-18).
- The Bath profile : an international Z39.50 specification for library applications and resource discovery. Release 1.1. Internationally registered profile / developed by The Bath Group; edited by The Bath Profile Editorial Team Carol Lunau, Paul Miller, and William Moen. Released June 2000. <http://www.ukoln.ac.uk/interop-focus/bath/current> (2000-11-07). Nova adresa profila: <http://www.nlc-bnc.ca/bath/bp-current.htm>.
- The Bath profile : what is it and why should I care? / prepared by Carol Lunau with input from the Bath Developer Group. <http://www.nlc-bnc.ca/bath/bathfaq.pdf> (2000-03-15).
- Berners-Lee, Tim. Why RDF model is different from the XML model. Last modified 1998-10-14. <http://www.w3.org/DesignIssues/RDF-XML.html> (2000-11-07).
- BiblioML : an XML application for bibliographic records, based on the UNIMARC Bibliographic Format, and for authority records, based on UNIMARC/Authorities / Ministère de la culture et de la communication, France, Mission de la recherche et de la technologie. <http://www.culture.fr/BiblioML/>
- Bosak, Jon. Media-independent publishing : four myths about XML. // IEEE Computer 31, 10(October 1998), 120-122. <http://metalab.unc.edu/pub/sun-info/xml/4myths.htm> (2001-01-26).
- CAIRNS Project recommendations for a cataloguing and indexing strategy for Scottish libraries. Glasgow : CAIRNS Cataloguing and Indexing Working Group, 2000. <http://cairns.lib.gla.ac.uk/docs/CAIRNSCatStrat.pdf>
- Cover, Robin. The XML cover pages. <http://www.oasis-open.org/cover/>
- Dubline Core Metadata Initiative. <http://purl.org/DC/>
- eXtensible Markup Language (XML) / W3C. <http://www.w3.org/XML/>
- HRN ISO 10163-1:1997 Informacije i dokumentacija - Međupovezivanje otvorenih sustava - Protokol za primjenu u službi pretraživanja i pronalazjenja - 1. dio: Specifikacija protokola (ISO 10163-1: 1993) = Information and documentation - Open systems interconnection - Search and retrieve application protocol specification - Part 1: Protocol specification (ISO 10163-1: 1993). 1. izd. [Zagreb] : Hrvatski zavod za normizaciju i mjeriteljstvo, 1997. [U postupku je zamjena ovoga standarda standardom HRN ISO 23950.]
- International Standard Maintenance Agency Z39.50 / The Library of Congress Network Development & MARC Standards Office. <http://lcweb.loc.gov/z39.50/agency/>
- ISO 2709:1981 Documentation - Format for bibliographic information interchange on magnetic tape. 2nd ed. [Genève] : ISO, 1981.
- ISO 8879:1986 Information processing - Text and office systems - Standard Generalized Markup Language (SGML). [Genève] : ISO, 1986.
- ISO 23950:1998 Information and documentation – Information retrieval (Z39.50) - Application service definition and protocol specification. 1st ed. [Genève] : ISO, 1998.
- Katić, Tinka. Z39.50 u sustavu OLIS (Oxford Library Information System). // 2. i 3. seminar Arhivi, knjižnice, muzeji : mogućnosti suradnje u okruženju globalne informacijske infrastrukture : zbornik radova / uredile Mirna Willer i Tinka Katić. Zagreb : Hrvatsko muzejsko društvo, 2000. Str. 139-148.
- Koch, Traugott. Practical experiences with interoperability and heterogeneity in digital library projects : Digital Cultural Heritage II : Interoperability of content, Amsterdam-Maastricht Summer

- University, Seminar. Maastricht, Netherlands 29 June - 1 July 2000. <http://www.lub.lu.se/tk/demos/Maastr0006.html>
- Lynch, Clifford A. The Z39.50 information retrieval standard : Part 1 : A strategic view of its past, present and future. // D-lib magazine (April 1997) <http://www.dlib.org/dlib/april97/04lynch.html> (1998-05-18).
- MacCallum, Sally. Extending MARC for bibliographic control in the web environment : challenges and alternatives. // Confronting the challenges on networked resources and the web / Bicentennial Conference on Bibliographic Control for the New Millennium, November 15-17, 2000 ; sponsored by the Library of Congress Directorate. http://lcweb.loc.gov/catdir/bibcontrol/mccallum_paper.html (2000-10-04).
- Miller, Dick R. XML : libraries' strategic opportunity. // Library journal (2000) <http://xmlmarc.stanford.edu/LJ/> (2000-10-18).
- Miller, Paul. Interoperability : what is it and why should I want it? // Ariadne 24(21-Jun-2000) <http://www.ariadne.ac.uk/issue24/interoperability/intro.html> (2000-10-18).
- Miller, Paul. Z39.50 za sve. // Vjesnik bibliotekara Hrvatske 43, 1-2(2000), 69-83.
- ONE-2 : OPAC Network in Europe. <http://www.one-2.org>
- Overview of SGML resources / W3C. <http://www.w3.org/MarkUp/SGML/>
- Resource Description Framework (RDF) / W3C. <http://www.w3.org/MarkUp/SGML/>
- Resource Description Framework (RDF) model and syntax specification / edited by Ola Lassila and Ralph Swick. <http://www.w3.org/TR/REC-rdf-syntax> (2001-01-27).
- Svenonius, Elaine. The intellectual foundation of information organization. Cambridge, Mass. : The MIT Press, 2000.
- TEI guidelines for electronic text encoding and interchange. <http://etext.virginia.edu/bin/tei-tocs?div=DIV1&id=SG> (2001-01-26).
- Willer, Mirna. UNIMARC u teoriji i praksi. Rijeka : Benja, 1996.