

POVEZANI PODACI I KNJIŽNICE

LINKED DATA AND LIBRARIES

Filip Farago

Hrvatska knjižnica i čitaonica "Đuro Sudeta", Garešnica
ffarago@ffos.hr

Boris Bosančić

Odsjek za informacijske znanosti, Filozofski fakultet Osijek
bbosancic@ffos.hr

Boris Badurina

Odsjek za informacijske znanosti, Filozofski fakultet Osijek
boris.badurina@ffos.hr

UDK / UDC 02:004

Pregledni rad / Review

Primljeno / Received: 23. 7. 2013.

Sažetak

U radu je dan prikaz koncepta povezanih podataka (*linked data*) kao semantičke nadogradnje postojeće mreže te razmotrena mogućnost integracije knjižničnih kataloga i usluga pomoću spomenutog koncepta u globalnu inteligentno povezanu mrežu – semantički web. Na početku rada govori se općenito o semantičkom webu i osnovnim načelima na kojima počiva. U dijelu rada koji razmatra koncept povezanih podataka, obrađene su njegove glavne sastavnice – URI (*Uniform Resource Identifiers*) i RDF (*Resource Description Framework*) – načela, koraci za objavu povezanih podataka na mreži te licenciranje i neka sigurnosna pitanja u vezi istih. U posebnom dijelu rada objašnjena je prednost pretraživanja povezanih podataka u odnosu na klasično pretraživanje mreže kao i problem primjene koncepta povezanih podataka na knjižničnim

zapisima. Analiziraju se svojstva MARC zapisa u mrežnom okruženju te razlika između povezanih podataka i tradicionalnih kataloga. Pokazano je da je knjižnični zapis u MARC formatu samo *čitljiv* dok bi isti takav zapis u RDF/XML obliku, integriran u povezane podatke, bio ne samo čitljiv, nego i stroju “razumljiv” u semantičkom smislu. Na kraju rada zaključeno je kako je tehnologija povezanih podataka velika prigoda za razvoj knjižničnih kataloga i usluga koje bi svoju ulogu u globalnoj inteligentno povezanoj mreži – semantički webu – ispunjavale na mnogo učinkovitiji način.

Ključne riječi: povezani podaci, URI, RDF, semantički web, knjižnice, MARC

Summary

The paper brings a review of the linked data concept as a semantic upgrade of the existing World Wide Web. It also analyzes the possibility of integrating library catalog and services into the intelligently linked network – semantic web. The first part of the paper presents some basic principles of the semantic web in general. Then we discuss the main elements of the linked data (URI - Uniform Resource Identifiers and RDF - Resource Description Framework), its principal concepts, publishing procedure, licensing and some security issues. The next section explains the main advantage of linked data searching in comparison with the conventional web search. Furthermore, we discuss the application of linked data on library catalogue records. MARC properties are analyzed in a network environment as well as the difference between catalogue with linked data and traditional library catalogue. The analysis has shown that MARC record is only readable, while the same record in RDF/XML format integrated into linked data is also understandable in the semantic sense. The conclusion is that the linked data technology presents a great opportunity for further development of library catalogues and library services, but also a was to improve efficiency in the intelligently linked global network.

Keywords: linked data, URI, RDF, semantic web, libraries, UNIMARC

Uvod

Globalni sustav hiperteksta i identifikatora dokumenta (*Uniform Resource Identifier – URI; Uniform Resource Locator – URL*), te pripadajući komunikacijski protokol (*Hypertext Transfer Protocol – HTTP*), koji danas nazivamo *web*, nastao je na temelju sustava koji je razvio Tim Berners-Lee krajem

1980-ih i početkom 1990-ih godina u CERN-u.¹ Njegov sustav razvijen je kao reakcija na probleme organizacije dokumenata u CERN-u: veliku dinamiku stvaranja i izmjene dokumenata, postojanje različitih, često nekompatibilnih formata pohrane, nekonzistentnost korištenja ključnih riječi, tj. indeksiranja unutar strukture stabla dokumenta koju je tamošnji sustav koristio i sl.²

U konceptualnim fazama razvoja mreže,³ oduvijek je bio naglasak na izgradnji sveopće platforme koja bi omogućila dijeljenje dokumenata na transparentan način. Na osnovnoj razini to je i postignuto jedinstvenim protokolom (HTTP), hipertekstom, tj. mehanizmom *hiperlinkova*, te sustavom URI / URL identifikatora. Premda je u izvornom dokumentu⁴ vidljivo da opisani sustav ima mogućnost semantičkog povezivanja, jer se govori o povezivanju čvorova bilo koje vrste na apstraktnoj razini, to još nije bilo praktično izvodivo na konzistentan način. Točnije, u navedenom prijedlogu iz 1989., Tim Berners-Lee predlaže, pored jednostavnih poveznica, model temeljen na označenim, usmjerenim grafovima u kojima linkovi i čvorovi imaju *tipove*.⁵ Navedeni koncept bit će tehnički formaliziran tek desetljeće kasnije u obliku RDF-a (*Resource Description Framework*).⁶

U prvoj javnoj verziji *weba* (web 1.0⁷) prevladavao je *top-down* model izgradnje statičnih stranica koje su bile namijenjene ručnom pregledavanju.

¹ Premda njegov sustav hiperteksta tada nije bio jedini, Tim Berners-Lee smatra se ocem mreže jer se njegov sustav razvio u globalnu mrežu zahvaljujući jedinstvenim prednostima u odnosu na postojeće zatvorene – vlasničke sustave. To su u prvom redu mogućnost jednosmjernog, kao i dvosmjernog povezivanja dokumenata pomoću hiperlinkova, te otvorenost protokola (koji nije vlasnički) što omogućuje jednostavnu sveopću – globalnu primjenu.

² Usp. Berners-Lee, Tim. Information management : a proposal. (March 1989, May 1990). [citirano: 2013-04-01]. Dostupno na: <http://www.w3.org/History/1989/proposal.html>

³ Isto.

⁴ Isto.

⁵ Isto.

⁶ Resource Description Framework (RDF) [citirano: 2013-04-02]. Dostupno na: <http://www.w3.org/RDF/>

⁷ Navedeni izraz *Web 1.0* nema službenu tehničku definiciju. Pod tim izrazom misli se na mrežu u počecima globalizacije interneta 1990-ih. Isto se odnosi na nazive *Web 2.0* i *Web 3.0*. Riječ je o “zvučnim” nazivima u popularnoj kulturi (engl. *buzzwords*). Pod pojmom *Web 2.0* općenito se smatra korištenje dinamičkih stranica i “semantičkih” društvenih mreža (npr., platforma za dijeljenje slika Flickr, pojava blogova, sadržaja koji stvaraju korisnici i sl.). Uz pojam *Web 3.0* vežu se napredne semantičke tehnologije povezivanja (koje bi trebale zaživjeti) – RDF, ontologije te koncept povezanih podataka (engl. *linked data*).

Ovakav pristup istovjetan je tradicionalnom pristupu koji koriste knjižnice – izradi MARC zapisa kao statičnih dokumenata. S rastom količine sadržaja, rastao je i problem njegove organizacije i pronalaženja. Za razliku od knjižnica, otvoreni *web* dopušta bilo kome stvaranje sadržaja i povezivanje s drugim sadržajima. Iako je ovakav liberalan pristup načelno dobar, poglavito sa stajališta slobodnog govora, sa stajališta organizacije, ovakav način povezivanja nema kvalitetu koju imaju dobro razrađeni kataložni standardi kakve koriste informacijske ustanove.

Ideja semantičkog weba, koju su 2001. predstavili Tim Berners-Lee i suradnici,⁸ uočava nedostatnost paradigme koja temelji informacijske sustave na pojedinačnim dokumentima namijenjenima ljudskoj obradi,⁹ te stavlja težište na apstraktniju razinu – problem predstavljanja znanja (engl. *knowledge representation*). U širem smislu, pod pojmom *semantičkog weba*, podrazumijevamo skup standarda W3 Konzorcija, krovnog tijela koje općenito donosi specifikacije, tj. preporuke vezane uz razvoj tehnologija *weba*,¹⁰ koje obuhvaćaju temeljne tehnologije: *Resource Description Framework* (RDF), *SPARQL Protocol* i *RDF Query Language* (SPARQL), *Web Ontology Language* (OWL), te *Simple Knowledge Organization System* (SKOS). W3 Konzorcij opisuje semantički web jednostavno kao mrežu podataka (engl. *web of data*)¹¹ koja se temelji na dva temeljna područja:¹²

- zajedničkim formatima za integraciju i kombiniranje podataka iz različitih izvora,
- jezicima kojima se opisuje kako se podaci odnose prema objektima iz stvarnog svijeta.

Povezane podatke (engl. *linked data*) Tim Berners-Lee predstavio je na TED konferenciji¹³ 2009. Riječ je o prirodnom razvoju semantičkog weba koji

⁸ Usp. Berners-Lee, Tim; James Hendler; Ora Lassila. The semantic web [citirano: 2013-04-01]. // Scientific American, 2001. Dostupno na: <http://cies.hhu.edu.cn/pweb/~zhuoming/teachings/MOD/N4/Readings/5.1-B1.pdf>

⁹ Ovaj problem jednako se odnosi na izvornu mrežu, kao i na metapodatkovne formate knjižnica (MARC).

¹⁰ Usp. World Wide Web Consortium (W3C) [citirano: 2013-04-01]. Dostupno na: <http://www.w3.org/>

¹¹ Usp. W3C semantic web activity [citirano: 2013-04-05]. Dostupno na: <http://www.w3.org/2001/sw/>

¹² Isto.

¹³ Usp. Berners-Lee, Tim. Tim Berners-Lee on the next web [citirano: 2013-04-01]. // TED Talks. / TED Conferences, LLC. Edinburgh, 2009.

težište stavlja na *podatke* i na globalno povezivanje skupova podataka (engl. *datasets*) koristeći RDF rječnike. *Library Linked Data Incubator Group*, koja djeluje pri W3 Konzorciju, koncept povezanih podataka opisuje kao podatke koji su objavljeni sukladno načelima ustrojenim tako da potiču povezivanje skupova podataka, skupova elemenata i rječničkih vrijednosti.¹⁴ Zajednica povezanih podataka¹⁵ – *Linked Data community* – opisuje povezane podatke kao skup dobrih praksi za objavljivanje i povezivanje strukturiranih podataka na mreži koristeći ključne tehnologije: URI, HTTP i RDF.¹⁶ U sklopu W3 Konzorcija, 2006. formirana je interesna skupina *The Semantic Web Education and Outreach* (SWEO)¹⁷ koja je djelovala dvije godine. Jedan od projekata koji je skupina začela bio je i *Linking Open Data : W3C SWEO Community Project*.¹⁸ Glavna svrha skupine je unapređenje zamisli semantičkog weba, tj. mreže podataka objavljivanjem raznih otvorenih skupova podataka povezanih RDF-om. Grafički prikaz *oblaka*¹⁹ povezanih podataka pokazuje porast broja skupova podataka odnosno ustanova uključenih u mrežu istih. U 2007. bilo je samo 12 povezanih skupova podataka, dok je njihov broj 2011. narastao na 295.²⁰

Općenito o semantičkom webu

Iako nema jedinstvene formalne definicije, općenito, semantički web je nadogradnja postojeće mreže u kojoj se informacijama dodjeljuje dobro definirano značenje koje omogućuje bolju suradnju računala i ljudi.²¹ Uza samu zamisao korištenja formalnih modela, bitan naglasak je na univerzalnosti koju

¹⁴ Usp. Library linked data incubator group final report [citirano: 2013-04-10]. Dostupno na: http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/#Scope_of_this_report

¹⁵ Usp. Linked data - connect distributed data across the web [citirano: 2013-04-10]. Dostupno na: <http://linkeddata.org/>

¹⁶ Usp. Linked data - connect distributed data across the web : frequently asked questions (FAQs) [citirano: 2013-04-10]. Dostupno na: <http://linkeddata.org/faq>

¹⁷ Usp. Semantic web education and outreach (SWEO) interest group [citirano: 2013-04-08]. Dostupno na: <http://www.w3.org/blog/SWEO/>

¹⁸ Usp. LinkingOpenData [citirano: 2013-04-08]. Dostupno na: <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

¹⁹ Usp. The linking open data cloud diagram [citirano: 2013-04-08]. Dostupno na: <http://richard.cyganiak.de/2007/10/lod/>

²⁰ Isto.

²¹ Usp. Berners-Lee, Tim; James Hendler; Ora Lassila. Nav. dj. Str. 1.

takve implementacije moraju zadovoljavati.²² W3 Konzorcij definirao je šest temeljnih načela na kojima se semantički web temelji. To su:²³

1. **Sve može biti identificirano URI-jem.** Sve objekte na mreži moguće je identificirati pomoću sustava identifikatora URI. Svatko tko ima nadzor nad dijelom mreže može dodjeljivati URI-je i određivati na što se oni odnose u fizičkom svijetu. URI-ji mogu (ali ne moraju nužno) biti ‘de-referencirani’ (engl. *de-referented*).²⁴ Rječnici kojima se opisuju izvori, također imaju URI-je.²⁵ URI (s URL-om i URN-om) predstavlja i najniži sloj postojećeg (‘predsemantičkog’) weba zajedno s Unicodom, univerzalnom kodnom tablicom koja se koristi za prikaz znakova pisama velike većine svjetskih jezika kao i za prikaz znakova pisama nekih drevnih jezika.²⁶
2. **Izvori i poveznice mogu imati tipove.** Današnja mreža, koja se sastoji od izvora i poveznica, namijenjena je ljudima za izravno korištenje. Takvi izvori nemaju metapodatke koji opisuju njihov kontekst i međusobne odnose. Primjerice, računalo ne može prepoznati je li neko djelo roman, ili je li neki dokument inačica nekoga drugog dokumenta. Izvori u semantičkom webu postaju strojno razumljivi koncepti. Digitalni objekti koji predstavljaju predmete iz stvarnog svijeta postaju “klase”, a odnosi među njima nadalje modeliraju se pomoću RDF-a.²⁷ Takvo formalno predstavljanje omogućuje prikazivanje podataka u obliku usmjerenog grafa.
3. **Djelomične informacije se dopuštaju.** Današnja mreža žrtvovala je integritet za skalabilnost, tj. stalan ili neočekivan rast. U njoj se dopuštaju poveznice koji ne vode nikamo. Možemo reći da se integritet pojedinih izvora postupno raspada. Semantički web također dopušta djelomične informacije, te je predviđeno da njegova pomagala mogu funkcionirati unatoč tome.

²² Isto. Korištena sintagma je “Anything can link to anything”.

²³ Usp. Koivunen, Marja-Riitta; Eric Miller. W3C semantic web activity [citirano: 2013-04-10]. Dostupno na: <http://www.w3.org/2001/12/semweb-fin/w3csw>

²⁴ Dereferencijacija URI-ja definira se kao preuzimanje (engl. *retrieving*) reprezentacije izvora koji URI identificira. Usp. Dereferencing HTTP URIs [citirano: 2013-04-12]. Dostupno na: <http://www.w3.org/2001/tag/doc/httpRange-14/2007-05-31/HttpRange-14>

²⁵ Primjerice, skup metapodataka DC (Dublin Core) za autora – *creator* ima URI <http://purl.org/dc/terms/creator>.

²⁶ Usp. Legg, Catherine. Ontologije na semantičkom webu. // Vjesnik bibliotekara Hrvatske 53, 1(2010), 160.

²⁷ Primjerice, RDF svojstvo `rdf:type` pojedini izvor može definirati kao instancu klase itd.

4. **Nema potrebe za apsolutnom istinom.** Na mreži ne postoji jedinstvena *normativna* baza podataka koja je provjereno istinita (niti je tako nešto moguće). Unutar semantičkog weba, svaka aplikacija zasebno zaključuje što je istina na osnovi dostupnih podataka. Pragmatično gledano, ona ne odlučuje o tome što je *istina*, nego kome će više *vjerovati*. Primjerice, podaci uglednih knjižnica mogu u nekom sustavu biti bolje rangirani od podataka neke manje poznate ‘treće strane’.
5. **Razvoj se podržava.** Različite skupine koje se bave nečime, mogu slično definirati neke koncepte. Pretpostavljen je stav da je okupljanje takvih sličnih koncepata iz različitih izvora korisno. Cilj je opisati izvore tako da dodavanje novijih podataka ne traži izmjenu starijih, nego da se naznače razlike i razriješe sukobi među njima. Potrebno je osigurati i da se konvencije mogu mijenjati kako razumijevanje nekog područja napreduje.
6. **Minimalističko oblikovanje.** W3 Konzorcij ima namjeru standardizirati samo najpotrebnije tehnologije. Zamišljeno je da se time olakšava izrada aplikacija na već postojećim dobro definiranim standardima, dok se istovremeno istražuju buduće složenije tehnologije.

Povezani podaci (*linked data*)

Općenito o povezanim podacima

U skladu s prethodnim izlaganjem, može se reći da koncept povezanih podataka predstavlja stvarnu, konkretnu implementaciju semantičkog weba ili kako se jednom prigodom izrazio i sâm Tim Berners-Lee: povezani podaci su zapravo “dobro napravljen semantički web” (engl. ‘*the semantic web done right*’).²⁸ Sa strogo tehničkog stajališta, kada se govori o konceptu povezanih podataka, misli se na metodu stvaranja poveznica između pojedinih ‘dijelova’ RDF dokumenata koji opisuju izvore. Kako bi to bilo moguće, svaki dio RDF dokumenta koji se želi povezati mora biti identificiran HTTP URI-jem. Može se reći da koncept povezanih podataka najviše počiva na dvjema glavnim tehnologijama semantičkog weba HTTP-u, URI-ju i RDF-u.

²⁸ Berners-Lee, Tim. The linked open data movement [citirano: 2013-04-01]. Dostupno na: <http://www.w3.org/2008/Talks/0617-lod-tbl/#%283%29>

URI

URI je najvažnija vrsta identifikatora za potrebe koncepta povezanih podataka. Njegova sintaksa određena je standardom RFC 3986²⁹ i definira se kao kompaktan niz znakova koji identificiraju apstraktni ili fizički izvor,³⁰ tj. niz znakova koji se podvrgavaju sintaktičkim pravilima³¹ i koriste set znakova *ASCII*.³² URI može koristiti razne *scheme*.³³ Radi se o prefiksima koji stoje na početku URI-ja. Za njihovu registraciju zadužena je agencija IANA - *The Internet Assigned Numbers Authority*.³⁴ Postoji pedesetak različitih shema. Neke od njih su *http*, *ftp*, *mailto*, *telnet*, *gopher* i druge. Za potrebe povezanih podataka, važna je jedino *http* shema. URI-ji temeljeni na njoj nazivaju se HTTP URI-ji. U sklopu specifikacije URI-ja, definira se i pojam izvora, važan za opis RDF-a. Specifikacija URI standarda definira izvor vrlo općenito. Izvor je bilo što, što se može identificirati URI-jem.³⁵ Bitno je naglasiti da URI jedinstveno identificira izvor, a ne reprezentaciju istog. Ako jedan izvor može imati više reprezentacija, na zahtjev klijenta, poslužitelj odabire odgovarajuću. To može biti dokument na jeziku korisnika, dokument prilagođen prikazu na mobilnim uređajima i sl.

RDF

Striktno gledano, RDF nije *jezik* u punom smislu te riječi. Sam RDF predstavlja *podatkovni model* koji se može izraziti kroz više strojno čitljivih jezika, a najčešći standard je RDF/XML.³⁶ RDF zajedno s *RDF Schemom* (RDFS) predstavlja temeljni sloj semantičkog weba sa stajališta razmjene metapodataka na podatkovnoj razini (s pretpostavljenom serijalizacijom u XML-

²⁹ Usp. Berners-Lee, Tim.; R. Fielding; L. Masinter. RFC 3986 - Uniform Resource Identifier (URI) : generic syntax. 2005. [2013-04-01]. Dostupno na: <http://tools.ietf.org/html/rfc3986>

³⁰ Usp. Isto.

³¹ Usp. Isto. Str. 4.

³² American Standard Code for Information Interchange (ASCII).

³³ Usp. Uniform Resource Identifier (URI) schemes [citirano: 2013-04-08]. Dostupno na: <http://www.iana.org/assignments/uri-schemes.html>

³⁴ Usp. The internet assigned numbers authority [citirano: 2013-04-01]. Dostupno na: <http://www.iana.org/>

³⁵ Usp. Berners-Lee, T.; R. Fielding; L. Masinter. Nav. dj. Str.4.

³⁶ Ostali strojno čitljivi jezici pomoću kojih se RDF model podataka može predstaviti u digitalnom okruženju su N-Triples, N3, Turtle, JSON.

u).³⁷ RDF izražava informacije pomoću jednostavnih *izjava* (engl. *RDF statements*). Njihova struktura sastoji se od tri dijela: subjekta, predikata i objekta. Takva izjava naziva se *trojka* (engl. *triple*). Na RDF grafu, subjekt i objekt predstavljaju čvorove (engl. *nodes*). Uzevši u obzir da URI-ji identificiraju bilo što kao izvor, subjekt i predikat RDF izjave *moraju* biti izvori prvog reda (tj. identificirani URI-jima), dok objekt RDF izjave može, ali ne mora biti izvor prvog reda odnosno može biti i običan niz znakova.³⁸ U praksi, to znači da ‘dijelovi’ RDF izjava koji su identificirani URI-jem mogu upućivati na ‘dijelove’ odnosno izvore drugih RDF izjava. Drugim riječima, objekt jedne RDF izjave može postati subjektom druge i tako se, barem u načelu, može ići unedogled. Opisani postupak, u najkraćim crtama, sažima bit koncepta povezanih podataka.

Konačno, skup međusobno povezanih RDF izjava čini jedan divovski globalni graf.³⁹

Načela povezanih podataka

U dokumentu u kojem opisuje temeljna načela povezanih podataka, Tim Berners-Lee definira četiri takva načela na kojima se grade povezani podaci – tzv. ‘*linked data principles*’:⁴⁰

1. Za imenovanje izvora odnosno ‘stvari’ (engl. *things*) koriste se URI-ji.
2. Potrebno je koristiti HTTP URI-je.
3. Kada netko pretražuje URI, potrebno je pružiti korisne informacije koristeći standarde semantičkog weba (RDF, SPARQL i sl.).
4. Potrebno je uključiti poveznice prema drugim URI-jima što omogućuje otkrivanje novih izvora odnosno *stvari* – informacija.

³⁷ RDF Scheme omogućuje standardizirani opis taksonomija i drugih jednostavnijih ontologija.

³⁸ Usp. Segaran, Toby; Jamie Taylor; Colin Evans. Nav. dj. Str. 65.

³⁹ Usp. Berners-Lee, Tim. Giant global graph. 2007. [citirano: 2013-04-01]. Dostupno na: <http://dig.csail.mit.edu/breadcrumbs/node/215>

⁴⁰ Usp. Berners-Lee, Tim. Linked data - design issues [citirano: 2013-04-01]. Dostupno na: <http://www.w3.org/DesignIssues/LinkedData.html>

Navedimo dodatno pojašnjenje za prvu stavku – imenovanje izvora URI-jima. Iako je sama tehnologija URI-ja jasno definirana standardom, postoje smjernice za oblikovanje istih. Postavio ih je opet Tim Berners-Lee, a URI-je oblikovane po njima – ‘*Cool URIs*’.⁴¹ Sažetak tih načela može se svrstati u tri kategorije.⁴²

- jednostavnost (engl. *simplicity*) – preporučljivi su kratki ‘mnemonički’ URI-ji koje je lako zapamtiti i slati elektroničkom poštom.
- stabilnost (engl. *stability*) – nakon što se za neki izvor postavi URI, poželjno je da tako i ostane, što je dulje moguće. Preporuča se izbjegavanje korištenja dijelova specifičnih za implementaciju, kao što je .php i sl.
- upravljivost (engl. *manageability*) – primjer dobre prakse je uključivanje godine u putanju, tako da možebitna promjena na URI-jima iz te godine ne utječe na ostale:⁴³

Iako postoji više shema identifikatora (kao što su URN i DOI), povezani podaci koriste samo HTTP URI-je. Dva su temeljna razloga za takvu odluku:⁴⁴

- HTTP URI-ji predstavljaju jednostavan način za stvaranje globalno jedinstvenih imena na decentralizirani način, s tim da svatko tko upravlja određenim imenskim prostorom može stvarati vlastite HTTP URI-je.
- HTTP URI-ji ne služe samo kao imena, nego i kao metoda pristupanja informacijama koje opisuju identificirani objekt.

⁴¹ Usp. Berners-Lee, Tim. Cool URIs don't change [citirano: 2013-04-01]. Dostupno na: <http://www.w3.org/Provider/Style/URI>

⁴² Usp. Cool URIs for the semantic web [citirano: 2013-04-04]. Dostupno na: <http://www.w3.org/TR/cooloris/#cooloris>

⁴³ Za HTTP 303 URI-je dobro je imati zasebnu poddomenu. Radi se o URI-ima koji identificiraju objekte iz stvarnog svijeta, a ne izvore na mreži. Poseban mehanizam preusmjerenja (engl. *redirection*) takve URI-je razrješava u mrežne dokumente koji opisuju te objekte iz stvarnog svijeta. Primjerice, na generički upit “Chimpanzee” koji identificira objekt iz stvarnog svijeta (životinjsku vrstu) server nema konkretan odgovor. Ovakvi upiti mogu se razriješiti upućivanjem na URI-je koji ih strojno opisuju. Primjer takvog projekta je <http://www.taxonconcept.org> iz područja biologije. Riječ je o taksonomiji vrsta u obliku povezanih podataka. Prethodni upit vodi na URI <http://lod.taxonconcept.org/ses/qgs5p>. Radi se o OWL ontologiji koja klasi čimpanza pridružuje čitav niz svojstava (poveznica na značajne izvore).

⁴⁴ Usp. Heath, Tom; Christian Bizer. Nav. dj. Str. 10.

Objava povezanih podataka na mreži

Objava povezanih podataka na mreži, sastoji se od tri općenita koraka:⁴⁵

1. Dodjeljivanje HTTP URI identifikatora entitetima koji se koriste za opis i osiguravanje mogućnosti *dereferencijacije* odnosno njihovih predstavljanja pomoću HTTP protokola.⁴⁶
2. Postavljanje RDF poveznica prema drugim izvorima podataka na mreži, kako bi se korisnici njihovim praćenjem mogli kretati mrežom podataka kao jednom cjelinom.
3. Potrebno je objaviti metapodatke o objavljenim metapodacima (u registrima metapodataka),⁴⁷ kako bi korisnici mogli procijeniti njihovu kvalitetu te izabrati mehanizam kojim će im pristupati.

Podaci ne moraju nužno biti i otvoreni i povezani. Naziv '*linked data*' odnosi se na tehničku interoperabilnost podataka, tj. na podatke koji su objavljeni sukladno navedenim načelima. Naziv *Open Data* odnosi se na legalnu (pravnu) interoperabilnost.⁴⁸

Vrlo korisni su i (neakademski) izvori koje stvaraju praktičari raznih profila okupljenih oko tehnologija semantičkog weba. Primjerice, Alex Williams⁴⁹ nudi četiri mjerila po kojima je potrebno vrednovati podatke kako bi odgovorili na pitanje radi li se o '*open linked data*':

- Podaci su javno dostupni i moguće ih je potvrditi.
- Podacima mogu pristupati i druge osobe i koristiti ih za vlastite svrhe.
- Podaci su povezani s drugim bazama podataka.
- U bazi podataka ugrađeno je znanje koje više ljudi može dijeliti.

Prednost povezanih podataka u odnosu na klasično pretraživanje mreže

Povezani podaci nisu nužno zamjena za tradicionalne tražilice, niti knjižnične kataloge, nego predstavljaju njihovu prirodnu nadogradnju. Također,

⁴⁵ Usp. Bizer, Christian; Tom Heath; Tim Berners-Lee. Nav. dj. Str. 5.

⁴⁶ U praksi vezanoj uz knjižnice, ovaj korak znači dodjeljivanje HTTP URI-ja nekom elementu bibliografskog opisa te objavljivanje tog URI-ja unutar nekog imenskog prostora. Tu funkciju obavlja OMR - *Open Metadata Registry* (<http://metadatatregistry.org/>), koji pomoću HTTP protokola omogućuje pristup URI-jima, njihovom opisu i RDF reprezentaciji.

⁴⁷ Vidi prethodnu napomenu.

⁴⁸ Usp. Library Linked Data Incubator Group Final Report. Nav. dj.

⁴⁹ Usp. Williams, Alex. The value of linked open data [Infographic] [citirano: 2013-04-06]. February 1st, 2011. Dostupno na: <http://readwrite.com/2011/02/01/the-value-of-linked-open-data>

oni ne predstavljaju nužno zamjenu za konceptualne modele bibliografskog opisa, kao što je FRBR.⁵⁰ Današnje tehnike pretraživanja koje koriste tražilice provode leksičku analizu dokumenata, njihovu matematičku obradu kao što je *Vector Space Model* (VSM) i algoritme koji su poslovna tajna. Tehnologija ovog tipa imat će svoje mjesto i u budućnosti. Međutim, ovakav pristup ima brojne konceptualne nedostatke. Primjerice, tražilice toga tipa ne mogu pretraživati podatke “kontekstualno”, tj. pretraživati skupove rezultata prema njihovim relacijama. Slabo strukturirani i/ili nepostojeći metapodaci, zajedno s hiperlinkovima na stranicama nisu dovoljni za strojno rukovanje rezultatima na takav složen način. Da bi korisnik dobio odgovor na upit koristeći konvencionalnu tražilicu (kao što je Google), mora mijenjati upite koristeći različite leksičke inačice pojmova. Ukoliko se traženi pojam ne nalazi u tekstu dokumenta, pretraživanje neće dati rezultat. Načelno, moguće je da dokument govori o nečemu, a što ne spominje izravno. Također, ovakav tip tražilice ne nudi mogućnost pretraživanja na jednom jeziku, a okupljanja dokumenata na drugim jezicima koji govore o traženoj temi. U tom scenariju, korisnik ima zadaću sâm istražiti postoje li srodni dokumenti na drugim jezicima. Prednost tehnologije povezanih podataka nalazi se upravo u takvim mogućnostima, a koje se provode na automatizirani način, ponajviše oslanjajući se i koristeći postojeće ontologije.⁵¹

⁵⁰ Konceptualne modele odnosa, kao što su FRBR, FRAD, FRASAD, moguće je zasebno implementirati (mapiranje obaviti na razini odnosa) u tehnologijama semantičkog weba.

⁵¹ Iako je riječ o izvorno filozofskom pojmu, u kontekstu semantičkog weba, pod pojmom *ontologije* smatra se složeni RDF rječnik čiji elementi (klase i svojstva) imaju jasno definirane tipove i *dvosmjerne* logičke veze koje omogućuju povezivanje s drugim klasama i svojstvima. S rastom semantičkog weba, raste i broj različitih ontologija koje opisuju pojedina područja. Navedimo samo neke primjere koji su značajni za primjenu unutar informacijskih ustanova: The Friend of a Friend (FOAF) ontologija (<http://www.foaf-project.org/>) koja se koristi za opis ljudi i relacija među njima, The Dublin Core (DC) ontologija (<http://dublincore.org/>) za općeniti metapodatkovni opis, Simple Knowledge Organization System (SKOS) ontologija (<http://www.w3.org/2004/02/skos/>) za strukturiranje kontroliranih rječnika pojmova, općenito predmetnica, te The Bibliographic Ontology (BIBO) (<http://bibliontology.com/>) za opis bibliografskih entiteta, klasifikaciju dokumenata koja se može koristiti i u svojstvu citatne ontologije. Važno je naglasiti da se ključne prednosti korištenja povezanih podataka ostvaruju upravo korištenjem ontologija. Zahvaljujući njima moguće je logičko zaključivanje, tj. izvođenje tranzitivnih relacija. Drugim riječima, nije potrebno izravno zadati sve odnose u samom skupu podataka, nego ih je moguće izvesti iz *aksioma* rječnika koji ih koristi. Usp. Bizer, Christian; Tom Heath; Tim Berners-Lee. Linked data - the story so far [citirano: 2013-04-04]. Dostupno na: <http://eprints.ecs.soton.ac.uk/21285/1/bizer-heath-berners-lee-ijswis-linked-data.pdf>, str. 59

Iako ovakav način povezivanja vjerojatno neće obuhvatiti cijelu mrežu, koja obuhvaća gomilu trivijalnih i besmislenih sadržaja, bit će od iznimne važnosti u svijetu znanosti i stručnih disciplina. Tipovi podataka koji će imati najviše koristi od uključivanja u semantički web, upravo su zapisi knjižnica, znanstvenih ustanova, zdravstva, vladinih podataka i sl. Radi se o podacima koji su dobro definirani, tj. imaju konzistentnu strukturu koja taj postupak olakšava. Ključna prednost tehnologije povezanih podataka u odnosu na današnje internetske tražilice leži u mogućnosti *netrivijalnih pretraživanja*. Točnije, koncept povezanih podataka omogućuje sustavu za pretraživanje, ne samo okupljanje zapisa oko nekog koncepta, nego i odgovaranje na pitanja koja prije toga još nisu bila postavljena. Moguće implementacije ovakvih sustava bile bi od velikog značaja u knjižnicama. Integracija takvih sustava u katalog, omogućila bi korisnicima dublji informacijski kontekst teme koju pretražuju, što je osobito korisno u slučaju istraživanja literature za potrebe znanstvenih i sličnih radova. Eksperimenti sa sustavima za preporučivanje glazbe, pokazuju da primjena povezanih podataka dovodi do bitnog povećanja *odziva i točnosti*.⁵²

Sigurnosna pitanja povezanih podataka

Pri razmatranju opsega primjene tehnologije *povezanih podataka*, valja uzeti u obzir da se radi o dodatnom naporu koji je potrebno uložiti u strukturiranje zapisa (mrežnih stranica, knjižničnih kataloga i sl.). Uzevši u obzir do sada poznatu praksu ugradnje proizvoljnih metapodataka u zaglavlja HTML dokumenata i zlouporabu istih (za lažno oglašavanje i sl., što je dovelo do toga da neke tražilice, npr., Google, te metapodatke gotovo u potpunosti ignoriraju), postavlja se pitanje legitimiteta koji će u ovoj tehnologiji zapisi imati. U načelu, u RDF izjavama, svatko može tvrditi bilo što. Koristeći mogućnosti složenog povezivanja i strojnog zaključivanja, moguće je zamisliti mnogo složenije obmane od jednostavnih primjera lažnog oglašavanja. U načelu, pitanje povjerenja – autentičnosti identiteta, koji se odnose na komunikaciju s nepoznatim pojedincima ili ustanovama preko mreže, na današnjoj mreži

⁵² Usp. Heitmann Benjamin; Hayes Conor. Using linked data to build open, collaborative recommender systems [citirano: 2013-04-10]. // AAAI spring symposium - technical report. SS-10-07: 2010. Dostupno na: www.aaai.org/ocs/index.php/SSS/SSS10/paper/download/1067/1452

rješava se digitalnim certifikatima. Oni se temelje na posredovanju treće strane koja je neutralna i kojoj obje strane vjeruju. Ova pitanja još nisu u središtu interesa zajednice koja se bavi povezanim podacima, jer tehnologija povezanih podataka još nije široko rasprostranjena.⁵³

Licenciranje povezanih podataka

Kako bi se ostvario jedan od temeljnih ciljeva otvorenih povezanih podataka – ponovna uporabljivost, postavlja se i pitanje njihova licenciranja. Među najvažnijima su normativne (engl. *authority*) datoteke. Knjižnice imaju dugu tradiciju u izradi i održavanju istih. Za očekivati je da su upravo normativne datoteke knjižnica dovoljno kvalitetne da služe kao osnova za neke daljnje primjene zainteresiranih trećih strana. Preporuke *Stanfordske radionice*⁵⁴ su stavljanje podataka u javnu domenu, tj. korištenje CC0 licence.⁵⁵ Ovu licencu usvojila je knjižnica *Europeana* za svoje metapodatke.⁵⁶ Primjere otvorenih licenci koje navodi Tim Berners-Lee⁵⁷ su *Creative Commons CC-BY* i *Open Government Licence* Ujedinjenog Kraljevstva. Međutim, valja znati da ne postoji globalni konsenzus oko pitanja autorskih i srodnih prava. Pravni sustavi različitih zemalja nisu još valjano definirali kako će se tretirati skupovi podataka. Vezano uz koncept *povezanih podataka*, problem je još složeniji. Uzevši

⁵³ S današnje perspektive, knjižnice su u dobrom položaju jer se radi o ustanovama koje imaju tradiciju izrade kvalitetnih zapisa. Velike knjižnice poput Kongresne knjižnice u Washingtonu svojim imenom jamče da njihovi zapisi objavljeni kao povezani podaci mogu služiti kao pouzdane reference u globalnim primjenama, za potrebe obrazovanja, kulture i sl. Jedan od mehanizama kojim je moguće 'brendirati' skupove podataka ili nadzirane rječnike pojmova, koji će biti objavljeni kao povezani podaci, leži u metodi oblikovanja URI-ja, tj. u korištenju imenskog prostora. Spomenuta Kongresna knjižnica svoje nadzirane rječnike naziva objavljuje unutar vlastitoga imenskog prostora koristeći se URI-jima zasnovanim na baznoj domeni (engl. base domain) "id.loc.gov". Svaka predmetnica (LCSH) mora imati HTTP URI i istu baznu domenu. Primjerice, koncept "Library science" posjeduje sljedeći HTTP URI: <http://id.loc.gov/authorities/subjects/sh85076723>.

⁵⁴ Usp. Be part of the web not just on it : report of the Stanford linked data workshop, 27 June - 1 July 2011. Washington, DC : Council on Library and Information Resources, 2011. [citirano: 2013-04-01]. Dostupno na: <http://www.clir.org/pubs/reports/pub152/LinkedDataWorkshop.pdf>, str. 28.

⁵⁵ Usp. CC0. Creative commons [citirano: 2013-04-04]. Dostupno na: <http://creativecommons.org/choose/zero/>

⁵⁶ Usp. Peters, Diane. Europeana adopts new data exchange agreement, all metadata to be published under CC0 [citirano: 2013-04-01]. Dostupno na: <http://creativecommons.org/weblog/entry/29133>

⁵⁷ Usp. Berners-Lee, Tim. Linked Data - Design Issues. Nav. dj.

u obzir njihov globalni karakter i strukturu, postavlja se pitanje primjenjivosti pojedinih zakona. Primjerice, američki zakon o zaštiti autorskih prava, štiti takozvana *kreativna djela*.⁵⁸ Pravno gledano, postavlja se pitanje smislenosti pozivanja na zaštitu autorskih prava kad su u pitanju baze podataka i njihovi elementi jer oni nisu *kreativno djelo*. Načelno, knjižnicama prihvaćanje otvorenih licenci ne bi trebao biti problem jer se radi o neprofitnim ustanovama. S druge strane, privatne ili komercijalne bibliografske agencije vjerojatno neće pristati na ovakav tip licenciranja.

Povezani podaci i kataložni zapisi

Semantički web i kataložni zapisi

Dramatičan rast mreže i njegov uspjeh možemo pripisati njegovoj univerzalnosti, tj. globalnom prihvaćanju zajedničkih otvorenih standarda. To omogućuje pretraživanje i pristupanje izvorima s bilo kojeg mjesta koristeći bilo koju modernu softversku platformu. Hipotetski, možemo pretpostaviti da mreža u takvom obliku ne bi postojala, niti opstala, da je isprva bila komercijalizirana na vlasničkim standardima. Analogno tome, uočavamo i problem bibliografskih formata knjižničnih zapisa. Pored međunarodnih razlika u izradi kataložnih zapisa na konceptualnoj razini (različiti kataložni pravilnici), arhaični formati pohrane i metode pretraživanja zapisa čine pojedine informacijske ustanove – otocima. Iako je moguće razmjenjivati zapise među pojedinim ustanovama, to nije moguće na transparentan način, kao što je to slučaj s javno dostupnom mrežom. Čak i u slučajevima kada je struktura formata zapisa (npr., MARC) međunarodno prihvaćena i otvorena, to se ne odnosi nužno i na njegovu implementaciju. Iako nigdje ne postoji zakon po kojem bi implementacija nekog standarda nužno morala biti otvorenog koda, poželjno je da je to tako.

Takva praksa, nepovoljna je iz više razloga:

- Zbog moguće nestandardne implementacije, otežava se ili onemogućuje suradnja (razmjena zapisa) s drugim knjižnicama u inozemstvu. Nasuprot tome, povezani podaci koriste globalno prihvaćenu otvorenu mrežnu arhitekturu.

⁵⁸ Usp. Miller, Paul; Rob Stylesb; Tom Heath. Open data commons, a license for open data [citirano: 2013-04-04]. // Populating the web with linked data : proceedings / the WWW2008 Workshop on linked data on the web, LDOW 2008, Beijing, China, April 22, 2008. Dostupno na: <http://ceur-ws.org/Vol-369/paper08.pdf>

- Odvojeni knjižnični sustavi nisu integrirani u otvorenu mrežu. Pretraživanje takvih zapisa nije na prikladan način dostupno neovisnim mrežnim tražilicama.
- Čitava knjižnična zajednica, ili njezin velik dio, ovisi o samovolji privatne korporacije.

Integracija katalognih zapisa u semantički web tražit će tehničke i neke konceptualne promjene u teoriji i načinu izrade zapisa. Sa stajališta konceptualnih promjena, neka arhaična katalogna pravila postat će manje važna, a moderna dostignuća, poput konceptualnih modela odnosa (FRBR i dr.) nastavit će svoj razvoj u novom okruženju semantičkog weba jer ih je moguće modelirati ontologijama. Discipline unutar knjižnično-informacijske znanosti koje se bave izradom normativnih datoteka te predmetnih odrednica, imaju iznimne mogućnosti unutar semantičkog weba.

Problemi primjene koncepta povezanih podataka na knjižničnim zapisima

Prema G. Dunsireu, glavni problemi objave knjižničnih bibliografskih podataka kao *povezanih podataka* mogu se podijeliti na nekoliko cjelina:⁵⁹

- **Nepostojanje RDF rječnika** – metapodaci koje koriste knjižnice nemaju svoju RDF reprezentaciju.
- **Nekompatibilnost** (s mrežnom tehnologijom) i manjkavost podataka. Značajne metapodatkovne komponente nisu identificirane URI-jima. Još ne postoji kritična masa knjižničnih bibliografskih zapisa objavljenih kao povezani podaci.
- **Organizacijski problemi**. Nedostatak otvorenih licenci koje omogućuju ponovnu uporabu povezanih podataka. Različite zajednice još nemaju stabilna *mapiranja* svojih metapodatkovnih shema s RDF-om. Ne postoji suglasnost oko optimalnog paketa ponovno upotrebljivih RDF metapodataka za aplikacije koje koriste povezane podatke knjižnica.
- **Tehnološki problemi**. Nedostatak otvorenih licenci koje bi obuhvatile API-je,⁶⁰ podatkovne standarde i klijentski softver u okruženju povezanih podataka.

⁵⁹ Usp. Dunsire, Gordon. Cluster BibData [citirano: 2013-04-04]. Dostupno na: http://www.w3.org/2005/Incubator/lld/wiki/Cluster_BibData

⁶⁰ API - *Application programming interface* (sučelje za programiranje aplikacija).

Sa stajališta semantike, za konverziju postojećih zapisa u *povezane podatke*, temeljni problem predstavlja odabir rječnika. Za strukturiranje i objavu nekog skupa podataka kao *povezanih podataka* postoje dvije mogućnosti: konstruirati vlastite rječnike (ontologije) ili koristiti postojeće.⁶¹ Općenito, u literaturi postoji suglasnost oko toga da je preporučljivo korištenje postojećih opće-prihvaćenih rječnika, gdje je to moguće.⁶² Sa stajališta kataloga koji koriste UNIMARC, problem je u tome što trenutno ne postoji reprezentacija istoga u RDF-u. To znači da još ne postoji imenski prostor u kojemu bi URI-ji za svaki element UNIMARC-a (RDF *klase* – subjekti) bili definirani, kao što ne postoji ni rječnik koji opisuje njihove odnose. Stoga, RDF reprezentacija kataloga temeljenog na UNIMARC-u, može imati dva glavna smjera:

- izradu cjelovite ontologije UNIMARC-a koristeći *RDF Scheme* (RDFS) ili OWL,
- povezivanje (mapiranje) pojedinih elemenata koristeći više različitih semantički najbližih rječnika.⁶³ U slučaju UNIMARC-a, najrodniji rječnik predstavlja ISBD⁶⁴ čija se RDF reprezentacija (imenski prostor) nalazi u registru RDF rječnika i elemenata metapodataka – Open Metadata Registry (OMR).⁶⁵

Trenutno, moguće je izvršiti konverziju UNIMARC zapisa u neki drugi format na sintaktičkoj razini bez gubljenja izvornih podataka. To nam omogućuje mehanizam XML Scheme koja omogućuje formaliziranje strukture XML

⁶¹ Knjižnična zajednica je do sad koristila relativno mali broj klasifikacijskih shema kojima se opisuju objekti iz stvarnog svijeta i bibliografski entiteti. Međutim, s razvojem semantičkog weba, broj RDF rječnika - ontologija, kojima se opisuju odnosi među entitetima, mjeri se u stotinama i vjerojatno će rasti u skoroj budućnosti. Integracija knjižnične katalogne infrastrukture u semantički web otvara nova zanimljiva pitanja. Pred knjižničnu zajednicu postavlja se novi izazov: kako klasificirati velik broj ontologija i njihovih elemenata, te time omogućiti knjižničarima i korisnicima pronalaženje odgovarajućih rječnika? Početkom 2013. godine, *Open Knowledge Foundation (OKFN)*, na svojoj mrežnoj stranici (<http://okfn.org/>) navodi 313 rječnika. Usp. *Linked open vocabularies (LOV)* [citirano: 2013-04-08]. Dostupno na: <http://lov.okfn.org/dataset/lov/>

⁶² Usp. Dunsire, Gordon; Mirna Willer. UNIMARC and linked data [citirano: 2013-04-10]. // *IFLA Journal* 37, 4(2011), 314-326. Dostupno na: http://www.ifla.org/files/assets/hq/publications/ifla-journal/ifla-journal-37-4_2011.pdf

⁶³ Usp. IFLA: ISBD/XML study group [citirano: 2013-04-10]. Dostupno na: <http://www.ifla.org/node/1795>

⁶⁴ Pored ISBD-a, radna skupina IFLA-e *ISBD/XML Study Group* ima zadaću usklađivanja imenskog prostora ISBD-a s drugima, što uključuje UNIMARC, kao i RDA/ONIX, DC - Dublin Core, a moguće i BIBO i FOAF ontologije.

⁶⁵ Preliminarnu reprezentaciju ISBD-a u RDF-u načinio je Gordon Dunsire. HTTP URI imenskog prostora glasi: <http://iflstandards.info/ns/isbd/elements/>

dokumenta.⁶⁶ Koliko je ovaj ‘međukorak’ koristan u postupku konverzije UNIMARC zapisa u RDF, vrijeme će pokazati.

Potreba za ovakvim pristupom leži u činjenici da su knjižnice baštinske ustanove koje pohranjuju građu na dugi rok. To se osobito odnosi na meta-podatke koje one izrađuju. Pozitivna strana takvog konzervativnog pristupa vidljiva je u primjeni obitelji MARC formata koji se uz manje izmjene stalno koriste već pola stoljeća. Sa stajališta primjene u knjižnicama, tehnologija povezanih podataka problematična je zbog toga što još ne postoje dobro utvrđena pravila, tj. primjeri iz prakse.

MARC zapis u mrežnom okruženju

Uočimo konceptualnu razliku u nazivlju koja se veže uza strojno čitljive knjižnične zapise. Format MARC (*MAchine-Readable Cataloging*) znači strojno čitljiv kataložni zapis. U vrijeme nastanka ovog formata, 60-ih godina prošlog stoljeća, pojam strojno čitljiv značio je samo to. Takav zapis bilo je moguće reproducirati koristeći tehnologiju onoga doba – računalom koje učitava podatke s magnetske trake i sl. Taj zapis ništa ne “radi”. On se učitava i prikaže. U tom smislu, MARC zapis samo je elektronička imitacija papirnata kataložnog listića. Kada danas govorimo o strojno čitljivom zapisu – u kontekstu semantičkog weba, pod strojno čitljivim smatramo više od samog čitanja i prikazivanja.

Tradicionalni kataložni zapisi – na listićima, kao i u MARC formatima, statični su i usmjereni prvotno tome da ih izravno koristi čovjek. Takvim zapisima nedostaje mogućnost interakcije s korisnikom i drugim zapisima. Koristeći format pohrane MARC, zapis postaje samo *čitljiv* u računalnom okruženju.

Današnji strojno čitljivi zapis, koristeći RDF/XML tehnologiju, nije samo čitljiv, nego i “razumljiv”. Koristeći formalizirane sheme kojima se pojedinim elementima zapisa – tagovima - daje ‘značenje’, dovodi do toga da je nad njima moguće strojno provoditi i složene logičke operacije. Zapise je moguće

⁶⁶ Zapis UNIMARC-a u XML formatu moguće je ostvariti korištenjem eksperimentalnih XML Schema za UNIMARC – BookMARC i UNIMARCSLIM. Usp. BookMARC: Prototype XML Schema for UNIMARC [citirano: 2013-04-04]. Dostupno na: <http://www.bookmarc.pt/unimarc/> Usp. UNIMARCSLIM: UNIMARC XML Schema prepared by Giovanni Bergamin and Detlev Schumacher for BNCf and based on MARCXML [citirano: 2013-03-30]. Dostupno na: <http://www.bncf.firenze.sbn.it/progetti/unimarc/slim/documentation/unimarcslim.html>

povezati s ontologijama u kojima su definirana svojstva i odnosi među objektima. Taj korak omogućuje računalu, ne samo “razumijevanje” zapisa, nego i rasuđivanje (engl. *reasoning*) o sadržaju i o odnosu s drugim sadržajem.

U klasičnom *online* katalogu, kao što je *CroList*, pretražujemo samo nizove znakova u nekom predefiniranom (UNIMARC) polju, a što može dovesti do toga da tražilica nije u mogućnosti pronaći izvor koji u bazi podataka ‘objektivno postoji’. Primjerice, pretraživanje autora upitom “*tatjana aparac*” ili “*tatjana aparac jelušić*” u katalogu knjižnice Filozofskog fakulteta u Osijeku⁶⁷ ne daje nikakve rezultate. Katalog naznačuje da se pretraživanje odvija po poljima “Odrednica / Uputnica”.

Metoda pretraživanja samo po UNIMARC nizovima znakova problematična je iz dva razloga:

1. Zbog pretraživanja po egzaktnom nizu znakova (koji korisnik ne može znati unaprijed), ne pronalaze se rezultati ukoliko ne postoji (kao što je to u ovom slučaju) softverski mehanizam koji bi po sličnosti povezo elemente upisa s normativnom bazom i ponudio najvjerojatnije tražene rezultate.
2. Knjižničar ne može očekivati od korisnika da unaprijed zna po kakvom je kataložnom pravilniku oblikovana osobna odrednica u UNIMARC polju 700, i to u njezinom točnom obliku “*Aparac-Jelušić*”.

Iz priloženog jasno je da izvor “*Aparac-Jelušić*” (kao entitet osoba) u UNIMARC-u ne postoji eksplicitno i to je osnovni problem ovog formata. UNIMARC poznaje samo oznake i zapise.⁶⁸ Lokalna ograničenost koja ga tjera da se oslanja isključivo na logiku kataložnog pravilnika, u najnovije vrijeme nastoji se prevladati uvođenjem novih potpolja koja onda UNIMARC zapisu omogućuju upućivanja na vanjske entitete.⁶⁹

Rasprave o modernizaciji MARC formata, zapravo MARC21 formata, korištenjem HTTP URI-ja, vežu se uz prvo desetljeće ovoga stoljeća. *Network*

⁶⁷ Usp. CROLIST: Sveuciliste J. J. Strossmayera u Osijeku, Knjižnica Filozofskog fakulteta [citirano: 2013-04-04]. Dostupno na: <http://crolist.ffos.hr/>

⁶⁸ Problem bi se djelomično mogao riješiti ‘intervencijom’ knjižničara kojemu UNIMARC nudi da u ponovljivom polju 400 unese neusvojene ili čak pogrešne oblike odrednice. Ako se ta polja dobro popune, a korisnik točno zna što traži, sustav će dati odgovor. Međutim, jasno je da ovo rješenje ovisi o knjižničaru, a ne o formatu pohrane.

⁶⁹ U ažuriranoj UNIMARC verziji iz 2012. polje 700 je dobilo potpolje \$o – međunarodni standardni identifikator naziva (*International standard identifier for the name*). Na ovaj način, u UNIMARC zapis moguće je uvesti jedinstveni entitet ‘definiran’ izvan njega. Usp. UNIMARC Bibliographic, 3rd edition updated [citirano: 2013-10-10]. Dostupno na: http://www.ifla.org/files/assets/uca/unimarc_updates/BIBLIOGRAPHIC/u-b_700_update.pdf

Development and MARC Standards Office Kongresne knjižnice u Washingtonu⁷⁰ te *MARC Advisory Committee* razmatraju uporabu URI-ja unutar MARC zapisa.⁷¹ Oni bi omogućili povezivanje pojedinih polja zapisa s nadziranim rječnicima pojmova, tj. koristili bi se kao identifikatori za pojedina polja zapisa. Nije usvojen konačan zaključak oko toga je li moguće koristiti URI-je kao identifikatore za povezivanje na navedeni način jer nije bilo jasno kako bi se oni tumačili. Jezikom RDF-a, nije riješeno što bi bilo subjekt, predikat i objekt. Isto tako, spomenuta rasprava nije razjasnila je li moguće ugrađivati URI-je u MARC zapis kao identifikatore polja na način na koji to čini RDF. Umjesto toga, korištenje URI-ja u potpoljima označenim s *\$u* može pomoći korisniku da sazna nešto više o navedenom predmetu odnosno vrijednosti polja klikom na ponudenu poveznicu.

Tehnički gledano, problem integracije MARC zapisa u semantički web leži u postojanju više inačica formata za koje još ne postoji suglasnost oko toga kako ih reprezentirati u RDF-u. Kako bi to postalo moguće, potrebno je izraditi poseban imenski prostor s pripadajućim RDF elementima i rječnicima koji bi opisivali sve elemente MARC zapisa i semantičke odnose među njima.

Razlika između povezanih podataka i tradicionalnog kataloga

Semantički povezani podaci omogućuju složena pretraživanja poput onih u relacijskim bazama podataka. To znači mogućnost postavljanja višestrukih logičkih uvjeta nad upitom, te probabilističko zaključivanje analizom *čvorova* u podatkovnom *oblaku*. U tehnologijama semantičkog weba, kataložni zapis mora postati u prvom redu podesan za strojno procesuiranje, tj. takav pristup je usmjeren na podatke. Izravna razumljivost korisniku temeljnih elemenata metapodatkovnog zapisa, sekundarna je. Korist koja nastaje iz takvih metapodataka, nužno je svojstvo koje proizlazi iz složenih interakcija metapodataka i (pretpostavljenog) bogatoga informacijskog metapodatkovnog krajolika.

Budući da tehnologija povezanih podataka omogućuje povezivanje različitih vrsta zapisa na mnogo načina, postavlja se legitimno pitanje postojanosti, preciznosti i prevelike zalihosti bibliografskog opisa podataka u takvoj tehnologiji. Kako bi osigurali postojanost pri izradi kataložnih zapisa, katalogizatori

⁷⁰ Usp. MARC 21 formats : guidelines for the use of field 856 [citirano: 2013-04-08]. Dostupno na: <http://www.loc.gov/marc/856guide.html>

⁷¹ Usp. MARC discussion paper no. 2009-DP01/1: Encoding URIs for controlled values in MARC records [citirano: 2013-03-20]. Dostupno na: <http://www.loc.gov/marc/marbi/2009/2009-dp01-1.html>

su kodificirali kataložne pravilnike u obliku velikog broja sintaktičkih pravila koja bi trebala pokrivati “sva” zamisliva stanja (npr., oblike imena, različite varijante i oblik istih za autore različitih nacionalnosti i sl.). Ovakav pristup ima velike praktične nedostatke. Nije realistično očekivati od različitih djelatnika da će ih ikad sva naučiti, potpuno razumjeti i/ili primjenjivati uvijek na isti način. Iz tog razloga, poželjno je korištenje zajedničkih, općeprihvaćenih globalnih baza (kataloga), kao što je VIAF. Nacionalne knjižnice, koje bilježe sve što je (službeno objavljeno) u zemlji, mogle bi imati značajnu ulogu u stvaranju normativnih datoteka za svoju zemlju i njihovoj objavi u obliku povezanih podataka, te povezivanju s ovakvim globalnim katalogima.⁷² Korištenjem međunarodnih baza podataka ovog tipa za sve elemente bibliografskog opisa, izrada kataložnih zapisa postala bi jednostavnijom. Neovisna izrada kataložnih zapisa, tj. udvostručivanje posla u svakoj knjižnici zasebno, postalo bi suvišno i nepoželjno. Konceptualno gledano, korištenje središnjih normativnih datoteka nije izum računalnog doba. Radi se samo o razvoju koncepta. Ključne prednosti korištenja jedinstvenih identifikatora za opis bibliografskih entiteta u modernom računalnom (mrežnom) okruženju, koje u “papirnom” dobu nisu postojale, jesu:

- Tehnička mogućnost stvaranja i korištenja takvih datoteka općenito (što ide u prilog ostvarivanju ideja univerzalnoga bibliografskog nadzora).
- Mogućnost korištenja istih u realnom vremenu (za neograničen broj korisnika).
- Čim postoji jedinstveni identifikator nekog “entiteta” u nekoj od normativnih datoteka, postojanje različitih oblika istoga, primjerice imena – nije problem. Odabir različitih varijanti, primjerice imena, postaje trivijalno pitanje koje se rješava na razini prikaza. To je omogućeno arhitekturom URI identifikatora. Za svaki *entitet*, dakle, postoji jedan identifikator i proizvoljno mnogo *reprezentacija* istog.⁷³

⁷² Pod ovom preporukom, misli se osobito na Nacionalnu i sveučilišnu knjižnicu u Zagrebu, koja u trenutku pisanja ovog rada, još nije na popisu ustanova suradnika VIAF-a.

⁷³ U stvarnosti nije realno za očekivati da će baš svaki entitet u informacijskom univerzumu posjedovati samo jedan identifikator. Prije bi se moglo reći da se radi o idealu kojemu većina normativnih baza stremi. Ipak, zahvaljujući svom dugogodišnjem iskustvu u radu na organizaciji i upravljanju nacionalnim bibliografijama, knjižnicama se pruža prigoda da svoje normativne baze ponude ‘tržištu’ semantičkog weba kao one koje bi su se navedenom idealu u budućnosti mogle najviše približiti.

Načelno gledano, knjižnična struka ima dobru teorijsku osnovu i veliko iskustvo u izradi klasifikacijskih shema (npr., Univerzalna decimalna klasifikacija) i bibliografskih modela (primjerice, FRBR,⁷⁴ FRAD,⁷⁵ FRASAD,⁷⁶ da spomenemo samo one novijeg datuma) u kojima se izriču problemi funkcionalnih zahtjeva za (normativne) bibliografske zapise, kao i primjerice ISBD,⁷⁷ koji se tiče elemenata opisa različitih vrsta građe. Stoga, barem načelno, integracija tehnologije *povezanih podataka* ne bi trebala predstavljati konceptualni problem za knjižničnu i srodne struke. Štoviše, korištenje *povezanih podataka* u baštinskim ustanovama logičan je slijed razvoja u području interoperabilnosti, jer implementacija tehnologija semantičkog weba ne zahtijeva nužno potpunu promjenu paradigmi u području katalogizacije, nego nudi tehničku metodu da se (korisna) teorijska dostignuća i konvencije na elegantan način uklope u globalno mrežno okruženje. Korisne osobine knjižničnih zapisa koji, barem načelno, olakšavaju njihovu integraciju u mrežno okruženje, jesu:

- Kvaliteta zapisa (u smislu konzistentnosti, kao i potpunosti opisa) – zapisi su rađeni po dobro razrađenim međunarodno prihvaćenim standardima.
- Brojnost zapisa – nakon usuglašavanja oko teorijskih i praktičnih problema vezanih uz konverziju knjižničnih zapisa u RDF odnosno *povezane podatke*, bit će moguće konvertirati velike količine zapisa istovremeno.
- Osiguravanju kvalitete povezivanja bibliografskih podataka u tehnologiji povezanih podataka uvelike će pridonijeti i korištenje velikih i kvalitetnih nadziranih rječnika pojmova te normativnih datoteka koje pokrivaju mnoga područja ljudskog znanja (primjerice VIAF, *Library of Congress Subject Headings – LCSH* i sl.). Pored njih, velike mogućnosti pridonnošenju porasta sveopće interoperabilnosti imaju i druge zajednice iz drugih, ne nužno knjižničnih, disciplina sa svojim rječnicima.

Stanfordska radionica⁷⁸ o povezanim podacima iznjedrila je suglasnost prisutnih oko sedam glavnih prednosti koje ima '*linked (open) data*' pristup.

⁷⁴ Functional Requirements for Bibliographic Records (FRBR).

⁷⁵ Functional Requirements for Authority Data (FRAD).

⁷⁶ Functional Requirements for Subject Authority Data (FRASAD).

⁷⁷ International Standard Bibliographic Description (ISBD).

⁷⁸ Usp. Be Part of the Web Not Just on It : Report of the Stanford Linked Data Workshop. Nav. dj. Str. 20.

To su:

1. *Linked open data* (LOD) stavlja informacije tamo gdje ih ljudi traže – na mrežu.
2. LOD poboljšava mogućnost otkrivanja (engl. *discoverability*) knjižničnih sadržaja.
3. LOD stvara mogućnosti za kreativne inovacije u sferi digitalnog učenja (engl. *digital scholarship*).
4. LOD omogućuje otvoreno i stalno poboljšavanje podataka.
5. LOD stvara bazu strojno obradivih (engl. *machine-actionable*) podataka na temelju kojih je moguće stvarati poboljšane servise.
6. Knjižnični povezani podaci mogu pomoći povezivanju različitih informacija iz područja.
7. LOD može pružiti izravan pristup podacima na način koji trenutno nije moguć, te biti koristan na načine kojih trenutno nismo svjesni.

Ukoliko knjižnice prijeđu s tradicionalnih bibliografskih formata na tehnologiju povezanih podataka, sam koncept izrade kataložnog zapisa doživjet će promjenu. Iako nije nužno mijenjati kataložne pravilnike – na neki način moguće ih je u novoj tehnologiji prevladati – velik dio tradicionalnih kataložnih pravila, prirodno će postati nepotrebnim. Pod tim se misli na niz kompliciranih pravila za uvrštavanje i oblikovanje osobnih imena, odrednica i sl. Danas, još uvijek, u manualnom papirnatom okruženju, (ali i u istovjetnim elektroničkim sustavima kao što je Crolist), izrada kataložnog zapisa postupak je koji generira statičan nepromjenjiv dokument. Transformacije knjižničnih zapisa u oblik namijenjen strojnom procesuiranju na razini značenja imat će za posljedicu mogućnost *razbijanja* tog postupka na više manjih i nezavisnih. Ukoliko se bibliografski elementi, tj. njihove vrijednosti pohranjuju pravilno u *atomiziranom* obliku,⁷⁹ tada je moguće korištenjem dobro definiranih shema odnosno ontologija pozivati bilo koji od njih i sukladno korisničkom zahtjevu, automatski ih smještati na odgovarajuće mjesto u bibliografskom zapisu. U tom slučaju, takav zapis se ne *izrađuje*, nego se *generira*. Na taj način, on može imati ‘proizvoljan’ sadržaj i opseg (uz pretpostavku da postoji pristup odgovarajućim elementima – varijantama imena na različitim jezicima i sl.). Razlog tomu leži u promjeni strukture zapisa. MARC zapis je statičan tekstualni dokument, dok su podaci u semantičkom webu strukturirani kao grafovi koje je poveznicama moguće neograničeno povezivati.

⁷⁹ Misli se na koncept *granularnosti* koji postoji u relacijskim bazama podataka.

Ono što je u tradicionalnom papirnatom katalogu na listićima bio jedan od temeljnih mehanizama organizacije – abecedni poredak, u mrežnom okruženju semantičkog weba ne znači mnogo. U katalogu na listićima, elementi abecede imaju ulogu operatora koji jednoznačno određuju poredak elemenata. Takav sustav može funkcionirati samo manualno. Točnije, unutarnja logika takvog sustava organizacije može funkcionirati u praksi samo onda kada postoji ljudski operater koji razumije te upute i postupa sukladno njima. Sa stajališta računala i strojne obrade podataka, abeceda ne predstavlja nikakvu samodostatnu logičku organizacijsku cjelinu, nego su njezini elementi, zajedno sa svim drugim znakovima, samo brojevi u nekoj kodnoj tablici. Ukoliko se (jezikom FRBR-a) za povezivanje entiteta i relacija među njima koristi strojno čitljivi RDF, koji povezuje *entitete* sustavom jedinstvenih identifikatora, utoliko korištenje abecednog poretka postaje trivijalna konvencija koja se može, a i ne mora koristiti pri prikazu.

U (pretpostavljeno bogatom) informacijskom okruženju povezanih podataka, čak i primjeri knjiga koje nemaju jasno definirane bibliografske podatke (namjerno nemaju naslov i sl.) nisu nužno problem za katalogizatore. Korištenjem identifikatora i ontologija, moguće je takav rad povezati s čitavim nizom različitih izvora (drugih kataloga, folksonomija i sl.) koji govore o takvom djelu i identificiraju ga u informacijskom krajoliku.

Zaključak

U radu smo razmotrili odnos koncepta povezanih podataka kao semantičke nadogradnje postojeće mreže i knjižničnih standarda na koje se oslanjaju tradicionalni katalozi. Izvorna mreža, iako ima dobre konceptualne strane, zapala je u organizacijsku krizu. Gomilanje sadržaja na mreži bez jasnih metoda organizacije stvara problem pri pretraživanju – gomilu beznačajnih rezultata. Stavljanjem u zajednički kontekst tehničke komponente mreže i knjižničnih bibliografskih zapisa, uočili smo da postoje zajednički problemi, kao i metode rješavanja istih. Primjerice, nadzirani rječnici i konceptualni modeli odnosa koje je razvila knjižnična zajednica, omogućili su točnija pretraživanja, ali nažalost, njihove mogućnosti nisu u potpunosti ostvarene. Razlog leži u tome što su kataložni zapisi temeljeni na obitelji MARC formata namijenjeni prvotno manualnom pregledavanju i korištenju. Njihov doseg ograničen je samo na lokalni katalog, a između pojedinih elemenata nije moguće uspostavljati logičke veze osim onih koje je predvidio katalogizator.

Tehnologija povezanih podataka velika je prigoda za integraciju knjižničnih kataloga i usluga u globalnu inteligentno povezanu mrežu – semantički web. Integracijom u istu, knjižnice, ne samo da mogu tehnički podići kvalitetu usluge, nego time mogu potvrditi predanost novoj paradigmi svog djelovanja u 21. stoljeću – širokoj integraciji u život zajednice. Iako je, naravno, ovaj cilj mnogo opsežniji od restrukturiranja samih kataloga, potonja zadaća od iznimne je važnosti. Uzevši u obzir pravno i financijski nametnute okolnosti nabave građe, koje su rezultirale promjenom politike – s posjedovanja k posredovanju, integracija knjižničnih izvora u semantički web, velika je prigoda da se ova zadaća ispuni na mnogo učinkovitiji način.

Važno je naglasiti da uočavamo razvoj pojma knjižnice. Nestaje koncepcija knjižnice kao statičnoga zatvorenog skladišta tiskane građe. Razlike između baštinskih ustanova pomalo se zamagljuju, te se otvaraju prostori suradnji kakva nekad nije bila moguća na jednostavan način. Umjesto *kataložnih zapisa*, govorimo o generiranju metapodataka. Katalog već sada možemo opisati kao *skupove podataka*. Takvo nazivlje usklađeno je s današnjim viđenjem šire zajednice baštinskih ustanova koje preispituju svoju teoriju i praksu, odnosno ulogu svojih ustanova u modernom umreženom društvu.

LITERATURA

Be part of the web not just on it : Report of the Stanford linked data workshop, 27 June - 1 July 2011. Washington, DC : Council on Library and Information Resources, 2011. [citirano: 2013-04-01]. Dostupno na: <http://www.clir.org/pubs/reports/pub152/LinkedDataWorkshop.pdf>

Berners-Lee, Tim. Cool URIs don't change [citirano: 2013-04-01]. Dostupno na: <http://www.w3.org/Provider/Style/URI>

Berners-Lee, Tim. Giant global graph. 2007. [citirano: 2013-04-01]. Dostupno na: <http://dig.csail.mit.edu/breadcrumbs/node/215>

Berners-Lee, Tim. Information management : a proposal. (March 1989, May 1990) [citirano: 2013-04-01]. Dostupno na: <http://www.w3.org/History/1989/proposal.html>

Berners-Lee, Tim. Linked data – design issues [citirano: 2013-04-01]. Dostupno na: <http://www.w3.org/DesignIssues/LinkedData.html>

Berners-Lee, Tim. The linked open data movement [citirano: 2013-04-01]. Dostupno na: <http://www.w3.org/2008/Talks/0617-lod-tbl/#%283%29>

Berners-Lee, Tim. Tim Berners-Lee on the next web [citirano: 2013-04-01]. // TED Talks. / TED Conferences, LLC. Edinburgh, 2009. Dostupno na: http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html

Berners-Lee, Tim; R. Fielding; L. Masinter. RFC 3986 – Uniform Resource Identifier (URI): generic syntax. 2005. [2013-04-01]. Dostupno na: <http://tools.ietf.org/html/rfc3986>

Berners-Lee, Tim; James Hendler; Ora Lassila. The semantic web [citirano: 2013-04-01]. // Scientific American, 2001. Dostupno na: <http://cies.hhu.edu.cn/pweb/~zhuoming/teachings/MOD/N4/Readings/5.1-B1.pdf>

Bizer, Christian; Tom Heath; Tim Berners-Lee. Linked data – the story so far [citirano: 2013-04-04]. Dostupno na: <http://eprints.ecs.soton.ac.uk/21285/1/bizer-heath-bernens-lee-ijswis-linked-data.pdf>

BookMARC: Prototype XML Schema for UNIMARC [citirano: 2013-04-04]. Dostupno na: <http://www.bookmarc.pt/unimarc/>

CC0. Creative commons [citirano: 2013-04-04]. Dostupno na: <http://creativecommons.org/choose/zero/>

Cool URIs for the semantic web [citirano: 2013-04-04]. Dostupno na: <http://www.w3.org/TR/cooluris/#cooluris>

CROLIST: Sveuciliste J. J. Strossmayera u Osijeku, Knjiznica Filozofskog fakulteta [citirano: 2013-04-04]. Dostupno na: <http://crolist.ffos.hr/>

Dereferencing HTTP URIs [citirano: 2013-04-12]. Dostupno na: <http://www.w3.org/2001/tag/doc/httpRange-14/2007-05-31/HttpRange-14>

Dunsire, Gordon. Cluster BibData [citirano: 2013-04-04]. Dostupno na: http://www.w3.org/2005/Incubator/lld/wiki/Cluster_BibData

Dunsire, Gordon; Mirna Willer. UNIMARC and linked data [citirano: 2013-04-10]. // IFLA Journal 37, 4(2011), 314-326. Dostupno na: http://www.ifla.org/files/assets/hq/publications/ifla-journal/ifla-journal-37-4_2011.pdf

Heitmann Benjamin; Hayes Conor. Using linked data to build open, collaborative recommender systems [citirano: 2013-04-10]. // AAAI spring symposium - technical report. SS-10-07: 2010. Dostupno na: www.aaai.org/ocs/index.php/SSS/SSS10/paper/download/1067/1452

IFLA: ISBD/XML study group [citirano: 2013-04-10]. Dostupno na: <http://www.ifla.org/node/1795>

Koivunen, Marja-Riitta; Eric Miller. W3C semantic web activity [citirano: 2013-04-10]. Dostupno na: <http://www.w3.org/2001/12/semweb-fin/w3csw>

Legg, Catherine. Ontologije na semantičkom webu. // Vjesnik bibliotekara Hrvatske 53, 1(2010), 155-206.

Library linked data incubator group final report [citirano: 2013-04-10]. Dostupno na: http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/#Scope_of_this_report

Linked data – connect distributed data across the web [citirano: 2013-04-10]. Dostupno na: <http://linkeddata.org/>

Linked data – connect distributed data across the web frequently asked questions (FAQs) [citirano: 2013-04-10]. Dostupno na: <http://linkeddata.org/faq>

LinkingOpenData [citirano: 2013-04-08]. Dostupno na: <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

Linked open vocabularies (LOV) [citirano: 2013-04-08]. Dostupno na: <http://lov.okfn.org/dataset/lov/>

MARC 21 formats guidelines for the use of field 856 [citirano: 2013-04-08]. Dostupno na: <http://www.loc.gov/marc/856guide.html>

MARC discussion paper no. 2009-DP01/1: Encoding URIs for controlled values in MARC records [citirano: 2013-03-20]. Dostupno na: <http://www.loc.gov/marc/marbi/2009/2009-dp01-1.html>

Miller, Paul; Rob Styles; Tom Heath. Open data commons, a license for open data [citirano: 2013-04-04]. // Populating the web with linked data : proceedings / the WWW2008 Workshop on linked data on the web, LDOW 2008, Beijing, China, April 22, 2008. Dostupno na: <http://ceur-ws.org/Vol-369/paper08.pdf>

Peters, Diane. Europeana adopts new data exchange agreement, all metadata to be published under CC0 [citirano: 2013-04-01]. Dostupno na: <http://creativecommons.org/weblog/entry/29133>

Resource Description Framework (RDF) [citirano: 2013-04-02]. Dostupno na: <http://www.w3.org/RDF/>

Segaran, Toby; Jamie Taylor; Colin Evans. Programming the semantic web. Beijing : O'Reilly, 2009.

Semantic web education and outreach (SWEO) interest group [citirano: 2013-04-08]. Dostupno na: <http://www.w3.org/blog/SWEO/>

The internet assigned numbers authority [citirano: 2013-04-01]. Dostupno na: <http://www.iana.org/>

The linking open data cloud diagram [citirano: 2013-04-08]. Dostupno na: <http://richard.cyganiak.de/2007/10/lod/>

Uniform Resource Identifier (URI) schemes [citirano: 2013-04-08]. Dostupno na: <http://www.iana.org/assignments/uri-schemes.html>

UNIMARCSLIM : UNIMARC XML Schema prepared by Giovanni Bergamin and Detlev Schumacher for BNCf and based on MARCXML [citirano: 2013-03-30]. Dostupno na: <http://www.bncf.firenze.sbn.it/progetti/unimarc/slim/documentation/unimarcslim.html>

W3C semantic web activity [citirano: 2013-04-05]. Dostupno na: <http://www.w3.org/2001/sw/>

Williams, Alex. The value of linked open data [Infographic] [citirano: 2013-04-06]. February 1st, 2011. Dostupno na: <http://readwrite.com/2011/02/01/the-value-of-linked-open-data>

World Wide Web Consortium (W3C) [citirano: 2013-04-01]. Dostupno na: <http://www.w3.org/>