Interoperabilnost i razmjena bibliografskih podataka

Cvitković Đone, Nina

Master's thesis / Diplomski rad

2018

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zadar / Sveučilište u Zadru**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:162:145720

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-28**

Image not found or type unknown

Repository / Repozitorij:

University of Zadar Institutional Repository

zir.nsk.hr

Image not found or type unknown

Sveučilište u Zadru

Odjel za informacijske znanosti
Diplomski sveučilišni studij Informacijske znanosti

Nina Cvitković Đone

Interoperabilnost i razmjena bibliografskih podataka

Diplomski rad

Zadar, 2018.

Sveučilište u Zadru

Odjel za informacijske znanosti
Diplomski sveučilišni studij Informacijske znanosti

Interoperabilnost i razmjena bibliografskih podataka

Diplomski rad

Student/ica:
Nina Cvitković Đone

Mentor/ica:
doc. dr. sc. Krešimir Zauder

Zadar, 2018.



Izjava o akademskoj čestitosti

Ja, **Nina Cvitković Đone,** ovime izjavljujem da je moj **diplomski** rad pod naslovom **Interoperabilnost i razmjena bibliografskih podataka** rezultat mojega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na izvore i radove navedene u bilješkama i popisu literature. Ni jedan dio mojega rada nije napisan na nedopušten način, odnosno nije prepisan iz necitiranih radova i ne krši bilo čija autorska prava.

Izjavljujem da ni jedan dio ovoga rada nije iskorišten u kojem drugom radu pri bilo kojoj drugoj visokoškolskoj, znanstvenoj, obrazovnoj ili inoj ustanovi.

Sadržaj mojega rada u potpunosti odgovara sadržaju obranjenoga i nakon obrane uređenoga rada.

Zadar, 19. rujna 2018.

Sadržaj

1 Uvod ... 1

2 Bibliografski zapis ... 3

2.1 Povijest bibliografskih zapisa ... 3

2.2 Bibliografski zapis i organizacija informacija ... 8

2.2.1 Bibliografski zapis i sadržajna obrada ... 11

2.3 Strukturirani zapis bibliografskih informacija ... 12

2.4 Izvori bibliografskih podataka: katalozi i baze ... 13

3 Web i faze u razvoju weba ... 18

3.1 Klasični web ... 19

3.2 Semantički web i umreženi podaci ... 20

3.3 Društveni web ... 24

3.4 Web podataka ... 26

3.5 Web i podaci ... 28

4 Interoperabilnost i razmjena podataka ... 30

5 Formati za razmjenu podataka ... 33

5.1 XML ... 34

5.2 JSON ... 36

5.3 Prednosti aktualnih formata za razmjenu ... 38

5.4 Model podataka ... 40

5.5 Umreženi podaci u JSON-u i XML-u ... 41

6 Razmjena bibliografskih podataka ... 43

6.1 Ograničenja u razmjeni i interoperabilnosti bibliografskih podataka ... 44

6.2 Bibliografski podaci i suvremeni formati za razmjenu ... 49

6.2.1 MARC XML ... 50

6.2.2 MODS 51
6.2.3 BibJSON 51
7 Zaključak: Dobra praksa u razmjeni bibliografskih metapodataka 53
8 Literatura 57

Sažetak

Rad istražuje mogućnosti interoperabilnosti i razmjene bibliografskih podataka pomoću aktualnih formata za razmjenu podataka. Cilj rada je objasniti važnost interoperabilnosti podataka s fokusom na bibliografske podatke. Opisat će se formati za razmjenu podataka koji koriste suvremene online informacijske usluge, XML i JSON, ali i formati za zapisivanje i razmjenu bibliografskih podataka, tj. MARC. Osim toga, opisati i analizirat će se i mogućnosti kombinacije tih formata s tradicionalnim knjižničnim standardima, kao što je MARCXML. Danas je bitno generirati različite oblike strukturiranih podataka za potrebe diseminacije i upravljanja kako bi se omogućili različiti vidovi korištenja tih podataka koji su potrebni korisnicima. Stoga će se analizirati tko koristi i razumije bibliografske podatke te prednosti suvremenih formata za razmjenu i zapisivanje podataka u različitim kontekstima njihova korištenja. Kroz rad će se objasniti i problem zapisa bibliografskih podataka iz MARC-a u neki od suvremenih oblika, problem složenih vrijednosti, pojava interpunkcije u samim vrijednostima (koje često dolaze od ISBD pravila). Također će se i definirati ključni pojmovi kao formati za razmjenu (XML, JSON, CSV), knjižnični standardi u tim formatima (MARCXML, MODS), aplikacijsko programsko sučelje (API), standardi, interoperabilnost podataka i ostali vezani koncepti.

Ključne riječi: Bibliografski podaci, razmjena podataka, interoperabilnost, formati za razmjenu podataka.

1 Uvod

U trenutku kada je količina građe u knjižnicama narasla toliko da je trebala pomoć u organizaciji došlo je do potrebe za bibliografskim zapisima. Bibliografski ili kataložni zapis je, prema Evi Veroni, "naziv za skup podataka kojima se u bibliotečnom katalogu, odnosno u bibliografiji opisuje određena jedinica bibliotečne građe i djelo sadržano u njoj. Izraz kataložni opis najčešće se upotrebljava za opis koji je izrađen za bibliografiju."[1]

Bibliografski zapis prenosi bibliografski opis koji je proizvod katalogizacije. Opis sadrži standardizirane elemente i podatke koji su potrebni kako bi se jedinica građe mogla jednoznačno identificirati. Katalogizacija se izvodi sistematski, precizno i opisuje bilo koje vrste građe koju knjižnica može prikupljati (rukopisna, tiskana, ili digitalna djela, razglednice, zvučni zapisi, filmovi i tako dalje). Standardi za opis postoje za svu vrstu građe koju knjižnice može posjedovati. Obzirom da služe identifikaciji publikacija, bibliografski zapisi su temelj izrade kataloga za pojedinu knjižnicu ili za izradu nacionalne bibliografije. Glavna svrha opisa je organizacija informacija odvojenih iz velike količine materijala, kako bi se olakšao pristup građi kao i korištenje samih informacija za potrebe upravljanja i istraživanja.[2] Bibliografski zapis se u katalogu može pretraživati prema autoru, naslovu, predmetu, ključnim riječima i slično. Kataloge mogu koristiti svi korisnici knjižnice.

S vremenom dolazi i do specijaliziranih koncepata poput bibliografske kontrole koju Joan M. Reitz definira kao “širi termin koji obuhvaća aktivnosti koje su uključene u stvaranje, organiziranje, upravljanje i održavanje bibliografskih zapisa, uključuje i standardizaciju opisa.”[3]

U suvremeno doba se sve više uviđa i potreba za razmjenom, a time i interoperabilnošću bibliografskih zapisa. Interoperabilnost prema definiciji Hrvatske enciklopedije “je sposobnost sustava, postrojbi ili snaga za pružanje i primanje usluga od drugih sustava, postrojbi ili snaga te uporaba tako razmijenjenih usluga za učinkovito međusobno djelovanje.”[4] Razmjena podataka je proces okupljanja podataka strukturiranih prema izvornoj shemi i interpretirajući

1 Verona, Eva. O kataložnom i bibliografskom opisu i njihovu ujednačavanju. // Vjesnik bibliotekara hrvatske 17, 1/2 (1971.). Citirano prema: Katić, Tinka. Stara knjiga: bibliografska organizacija informacija. Zagreb: Hrvatsko knjižničarsko društvo, 2007. Str. 60-61.
2 Usp. Bibliography. Encyclopædia Britannica. URL: https://www.britannica.com/topic/bibliography (2018-05-02)
3 Reitz, Joan M. Bibliographic database. // Dictionary for Library and Information Science. Westport, Connecticut: Libraries Unlimited, 2004. Str. 69.
4 Interoperabilnost. // Hrvatska enciklopedija. URL: http://www.enciklopedija.hr/natuknica.aspx?id=27658 (2018-05-18)

ih u podatke strukturirane za ciljanu shemu, tako da interpretirani podaci predstavljaju točan prikaz izvornih podataka.[5]

Iako knjižničarski sustavi u načelu posjeduju sve elemente interoperabilnosti i razmjene iz definicija, World Wide Web je znatno nadmašio i iskoristio mogućnosti za interoperabilnost i razmjenu informacija. Tijekom posljednja dva desetljeća, web je rapidno napredovao: ne samo da su umreženi dokumenti već i podaci. Danas korisnici mogu pristupiti strukturiranim informacijama u svakom trenutku te ih na različite načine pobirati, organizirati i razmjenjivati. Kao glavni izvor informacija smatraju web te su formati koji se koriste za zapisivanje, razmjenu i interoperabilnost na webu postali *de facto* standard za strukturirane podatke. Knjižničarski standardi i sustavi posustaju i ne mogu pratiti promijene koje donosi web. Ne prateći nove trendove i mogućnosti dovodi knjižnice u opasnost te nisu u mogućnosti zadovoljiti potrebe suvremenih korisnika. Ne samo da današnji korisnici zahtijevaju drugačiji pristup organizaciji informacija, već bi i knjižničarska zajednica imala koristi od novih ideja i mogućnosti koje donosi web i umreženi podaci jer se mnoge vezane teme tiču upravo organizacije informacija.

[5] Usp. Doan, AnHai. Halevy, Alon. Ives, Zachary. Principles of Data Integration. Waltham, MA: Morgan Kaufmann, 2012. Str. 273. URL: https://books.google.hr/books?id=s2YCKGrO10YC&printsec=frontcover&redir_esc=y#v=onepage&q&f=false (2018-05-20)

2 Bibliografski zapis

2.1 Povijest bibliografskih zapisa

Iako se ne zna točna godina kada se pojavila prva knjižnica, naznake se mogu pronaći još u Sumeru čak unazad 2600. godina prije nove ere. U počecima su knjižnice bile spremišta za glinene pločice ispisane klinastim pismom, često religijskog sadržaja.[6] Od samog početka nije postojala potreba za bibliografskim zapisima, ali u trenutku kada su kolekcije postale prevelike da bi pojedinci mogli jednostavno pristupiti i pronaći ono što traže, knjižničari su morali osmisliti neku vrstu „surogata“ koja bi reprezentirala sve primjerke u knjižnici i organizirala ih. Zapravo je bibliografski opis usko vezan uz razvoj i napredak knjižnica i knjižničarstva.

Prve oblike bibliografskog zapisa možemo pronaći već u 7. stoljeću prije nove ere u Asurbanipalovoj knjižnici koja je posjedovala 30 000 glinenih pločica na nekoliko jezika, organiziranih prema obliku i sadržaju. Naslovi i imena autora su postavljani na zid kako bi korisniku olakšali pronalazak. Slijedili su Babilonska pravila te je bio običaj organizirati građu prema sadržaju. Praksa da se građa organizira prema sadržaju je bila opće poznata u to doba pošto se mogu uočiti pravila prema kojima su knjižničari relativno ujednačeno označavali građu tako da su korisnici znali kako pretraživati kada dođu u knjižnicu.[7] Glavni fokus tadašnje organizacije i zapisa je bio sadržaj djela iako su zapisivali ime i prezime autora i naslov. Uglavnom se izabralo par glavnih sadržajnih tema prema kojima je građa bila organizirana. Takva praksa organizacije se nastavlja i mnogo stoljeća kasnije.

U staroj Grčkoj bibliografski zapisi su bilježeni na drvenim pločama koje su se nalazile iznad kutija sa svitcima. Kao ocem bibliografije i bibliografskih zapisa, smatra se Kalima, starogrčki pjesnik, kritičar i učenjak. On je jedan od najznačajnijih knjižničara Aleksandrijske knjižnice iako nigdje nije službeno pronađeno da je radio kao knjižničar. Pretpostavlja se da je tada knjižnica posjedovala oko 400,000 papirus svitaka koji su bili grupirani prema sadržaju. Kalima je napisao djelo koje je predstavljalo sva djela koje je knjižnica posjedovala, danas poznato pod imenom “Pinakes”. Pinakes je bio 120 svitaka dug, a danas postoje samo fragmenti. Na njima su zapisane oznake koje su se nalazile iznad kutija u kojima su bili svisci. Analizirajući Kalimaove Pinakes nije sigurno da li je to, u današnjem smislu, bio katalog ili

[6] Usp. Harris, Michael H. History of libraries in the western world. 4. izd. Metuchen: Scarecrow Press, 1995. Str. 3-17.
[7] Usp. Harris, Michael H. History of libraries in the western world. 4. izd. Metuchen: Scarecrow Press, 1995. Str. 17-27.

bibliografska biografija. Dilema postoji pošto se osim naslova i autorova imena (elementi koji se i danas zapisuju) moglo pronaći i autorovo mjesto rođenja, ime oca, ponekad i učitelja te informacije o autorovu obrazovanju. Osim toga, opisan je i ukratko autorov životopis, popis njegovih djela i koliko svako djelo ima linija teksta, prva linija teksta o kojem je bibliografski zapis (u slučaju poezije prvi stih) i informacija o podrijetlu svitka. Značajno kod Kalima je i to što je prvi podijelio Pinakes u 8 glavnih kategorija prema sadržaju. Pločice su bile kronološki i u cijelosti uređeni katalog knjižnice te su kasnije korištene kao temelj za radove na temu povijesti grčke književnosti. Kalimaov popis je značajan za povijest bibliografskih zapisa i kataloga, iako ne postoje dokazi da se popis stvarno koristio kao katalog.[8]

Kalimaov bibliografski opis je, uspoređujući ga s današnjim, kombinacija bibliografskog opisa i bibliografije. Iako danas možda ne vidimo potrebu za svim elementima koje je on koristio, sigurno je da je opis bio prikladan za vrijeme u kojem je bio napisan i tadašnje zahtjeve korisnika. Samo postojanje "kataloga" ukazuje na svijest o potrebi za okupljanje informacija o građi koju knjižnica posjeduje kao i potrebu za odvajanje metapodatkovnih zapisa te upravljanje i pretraživanje putem metapodatkovnih surogata.

U početcima su se opisi nalazili iznad građe na pločicama ili svitku koji nisu bili podložni pretraživanju kao što je to i danas slučaj kod fizičkih objekata (bilo da se radi o tiskanoj građi, medijima za pohranu analognih ili digitalnih zapisa ili pak čemu trećem). Naravno ne smije se zaboraviti da su zbirke bile izrazito manje, npr. u srednjem vijeku su knjižnice posjedovale od oko 300 do 650 volumena. Pošto su knjižnice najčešće bile u sklopu nekog samostana, težilo se da knjižnica posjeduje jedan primjerak za svakog redovnika, a onda kada bi se taj kriterij ispunio zbirka bi sporije nastavila rasti. Redovnici koji su bili odgovorni za knjižnice su uvijek imali još dodatnih zadataka pošto im posao u knjižnici nije iziskivao toliko vremena. U 7. stoljeću, Izidor Seviljski je organizirao knjižničnu zbirku na 14 polica gdje je svaka polica bila posvećena jednom autoru. Iznad svake police je bila pjesma o autoru i za svaku skupinu je Isidor napisao opisnu pjesmicu.[9]

U početcima, knjižnice su svoje zbirke uglavnom organizirale prema sadržaju, a ponekad prema veličini ili prema tome kako se građa nabavljala. Ovakva praksa se zadržava i kasnije kada knjižnice kreću rasti. Građa se i dalje raspoređivala prema gruboj sadržajnoj podjeli te na kraju svake police je bio popis knjiga koje su se nalazile na njoj. Tada su se koristile lokalne

[8] Usp. Harris, Michael H. History of libraries in the western world. 4. izd. Metuchen: Scarecrow Press, 1995. Str. 44-46.
[9] Isto. Str. 93-101.

klasifikacije koje su pružale simbole za lokaciju, simboli su se referirali na područje i policu na kojoj se knjiga nalazila te bi svaka knjiga dobila svoj broj. Ideja o univerzalnoj klasifikaciji koju bi dijelilo više knjižnica još nije postojala, kao ni ideja o standardizaciji za bibliografski opis. Poznato je i da je jedna knjižnica svoju inventarnu knjigu organizirala prema prvoj riječi koja se nalazila na drugoj stranici knjige.[10] Ipak, čak i tijekom modernog razdoblja nije postojala univerzalna metoda za organizaciju već je odluke donosio direktor knjižnice.

Tek pojavom tiskarskog stroja i povećanjem broja knjiga je došlo do potrebe za standardizacijom organizacije informacija. U 18. stoljeću nastaju francuska pravila katalogizacije. U pravilima katalogizacije iz 1791. godine je propisano da se naslovna stranica treba prepisati na karticu i autorovo prezime treba biti podcrtano, a u slučaju da nema autora podcrta se ključna riječ u naslovu te se dodaje oznaka za razvrstavanje u koju se uključivalo broj volumena, veličina, podatak o ilustraciji, materijalu, vrsta formata, informacije o nedostajućim stranicama i opis uveza ako je potrebno. Nakon što bi se kartice ispunile i organizirale prema podcrtanim riječima uvezane su zajedno u donjem lijevom kutu.[11] Treba se istaknuti i godina 1826. kada je Američko Filozofsko društvo predložilo da bi katalog trebao služiti kao korisnički alat, a nakon što je ovo postalo pravilo, potreba za standardizacijom je postala neophodna.[12]

Klasičnu definiciju svrhe knjižničnog kataloga dao je Charles A. Cutter, knjižničar iz 19.st., koji je naveo ove ciljeve kataloga:

1) „omogućiti osobi da pronađe knjigu o kojoj zna bilo koji od sljedećih elemenata:
   a) autora, b) naslov, c) predmet
2) pokazati što knjižnica posjeduje:
   a) od određenog autora, b) o određenoj temi, c) od određenog književnog roda (pjesništvo, drama, proza)
3) pomoći u izboru djela:
   a) s obzirom na izdanje (bibliografski), b) s obzirom na vrstu (književna ili predmetna)“[13]

[10] Isto. Str. 114.
[11] Taylor, Arlene G; Jourdey, Daniel N. The Organization of Information. 3. izd. Westport, Connecticut: Libraries Unlimited, 2009. Str. 72-73.
[12] Usp. Clarke, Rachel Ivy. Breaking Records: The History of Bibliographic Records and Their Influence in Conceptualizing Bibliographic Data. // Cataloging & Classification Quarterly 53, 3-4 (2014.). Str. 288.
[13] Horvat, Aleksandra. Knjižnični katalog i autorstvo. Rijeka: Naklada Benja, 1995. Str. 49.

Cuttlerova pravila za katalog su pružila naznake koje su informacije važne za kataložni zapis. Skupine podataka, iako slične, koje su se zapisivale u bibliografskim opisima, u bliskoj prošlosti možemo podijeliti u nekoliko era. Kao prva, može se istaknuti pravila za Bodleianu iz 1674. godine koja su propisivala sljedeće elemente: autorovo prezime, naslov (ukratko) i broj i veličina volumena, ako je potrebno datum i mjesto publikacije i napomena za identitet donatora. Sljedeća poznata pravila su Pazinijeva 91. pravila iz 1841. godine, za koja se smatraju da su temelji današnjih. Pazinijev opis se sastojao od: naslova, autora, podatak o ne savršenstvu na knjizi, broj volumena i dijelova, mjesto izdanja, nakladnikovo ime, datum izdanja i napomene o primjerku (da li je knjiga tiskana na finom papiru i slično). Do 1980-ih kao ključni elementi zapisa su izdvojeni: naslov, podatak o odgovornosti, izdanje, podaci o izdavanju, materijalni opis, nakladnička cjelina, napomene i standardizirani identifikatori poput ISBN-a.[14] Danas podjelu ovih elemenata poznajemo pod nazivom ISBD (International Standard Bibliographic Description ili Međunarodni standardni bibliografski opis). Tek se u dvadesetom stoljeću donosi široko prihvaćeni standardi za opis građe i klasifikaciju sadržaja, čime će se omogućiti razmjena opisa među katalozima.

Eva Verona ističe važan zaokret u kataložnim pravilima kada se odustalo od stvaranja međunarodnih kataložnih pravila, već se usmjerilo na ujednačavanje kataložnih načela prema kojima je trebalo ujediniti nove nacionalne kataložne pravilnike. Na zasjedanju IFLA-e 1961. u Parizu prihvaćena su tzv. Pariška načela kojima su određeni jedinstveni postupci u izradbi odrednice za svaku publikaciju, uključujući u pojam publikacije sve vrste knjižnične građe. Eva Verona, koja je djelatno sudjelovala u radu Komisije za izradbu pravilnika, objavila je 1970. godine, u izdanju Hrvatskoga bibliotekarskog društva iz Zagreba, Pravilnik i priručnik za izradbu abecednih kataloga.[15]

Do 20. stoljeća nije postojalo univerzalno pravilo koje bi određivalo koje se informacije trebaju zapisivati za bibliografski opis, što je i logično obzirom da različite vrste građe i zbirki te različite skupine korisnika imaju potrebe za različite elemente bibliografskih opisa. Ipak kroz povijest se definiraju neki standardni elementi bibliografskog opisa poput naslova, autora i ostalih podataka o odgovornosti, vremena objave (najčešće godina). Porastom količine i vrsta

[14] Usp. Clarke, Rachel Ivy. Breaking Records: The History of Bibliographic Records and Their Influence in Conceptualizing Bibliographic Data // Cataloging & Classification Quarterly 53, 3-4 (2014.). Str. 286-302.
[15] Usp. Katić, Tinka. Stara knjiga: bibliografska organizacija informacija. Zagreb: Hrvatsko knjižničarsko društvo, 2007. Str. 61.

publikacija, bibliografski opis postaje sve složeniji kako bi zadovoljio potrebe jednoznačne identifikacije i pronalaska građe.

Današnji oblici bibliografskih zapisa potječu iz vremena tradicionalnih knjižnica. Nakon Pariških načela iz 1960-tih, 1969. Godine u Kopenhagenu se prihvaća prijedlog da se na temelju studije M. Gormana izradi nacrt za standardni bibliografski opis. Iz ovog nacrta nastao je *International Standard Bibliographic Description* (ISBD) ili *Međunarodni standardni bibliografski opis*. Postojala je potreba za standardom koji bi propisivao izgled i sadržaj bibliografskog opisa. Prvo je objavljen *Međunarodni standardni bibliografski opis za omeđene publikacije* (ISBD(M)), 1971. godine. Ovaj standard ubrzo prihvaćaju Europske nacionalne knjižnice, a 1974. godine je objavljen standard za serijske publikacije. Zbog postojanja različite vrste građe koje su knjižnice posjedovale, ubrzo je uviđena potreba za ISBD-om koji bi bio prikladan za sve vrste građe ili ISBD(G), a nakon toga su objavljeni ISBD-ovi za svaku pojedinu vrstu.[16] Godine 2011. je objavljeno konsolidirano ili objedinjeno izdanje.

Osim ISBD-a, za razvoj bibliografskih opisa također je bitan i razvoj MARC formata i to posebno u kontekstu razmjene zapisa. Strojno čitljivo katalogiziranje (Machine-Readable Cataloging) „ je skupina formata koji primjenjuju određeni skup konvencija za identifikaciju i uređivanje bibliografskih podataka radi obrade računalom."[17] Vrlo važna značajka MARC-a je da MARC zapisi nisu namijenjeni krajnjim korisnicima već služe internoj organizaciji, a na temelju tih zapisa se automatski izrađuju različiti oblici podobni za korisnike. Ova ideja, kao što ćemo vidjeti kasnije, postaje vrlo značajna u suvremenom digitalnom umreženom okruženju.

Danas se mogu odvojiti dvije glavne verzije MARC-a, MARC 21 i UNIMARC. Razlika između ove dvije verzije MARC-a se odnosi na same nazive polja i potpolja, a u suštini dijele pristup organizaciji informacija. Prva verzija MARC-a razvija se već 1960-tih godina, a do 1971. godine MARC postaje standard za bibliografske podatke u Sjedinjenim Američkim Državama. Zapise je radila Kongresna knjižnica koja je centralna knjižnica u SAD-u. Ovakva vrsta katalogizacije ne samo da je ujednačila kataložne zapise, već je i smanjila troškove, osigurala i olakšala pristup informacijama korisnicima i knjižnicama. Kasnije, spajanjem

[16] Usp. ISBD: međunarodni standardni bibliografski opis / preporučila Skupina za pregled ISBD-a; odobrio Stalni odbor IFLA-ine Sekcije za katalogizaciju; [s engleskog prevela i predgovor napisala Ana Barbarić]. Objedinjeno izd. Zagreb: Hrvatsko knjižničarsko društvo, 2014. Str. XVIII.

[17] Gradley, E. Hopkinson, A. Exchanging bibliographic data: MARC and other international formats. Ottawa [etc..]: Canadian Library Association [etc…], 1990. Str. 70. Citirano prema: Willer, Mirna. UNIMARC u teoriji i praksi. Rijeka: Naklada Benja, 1996. Str. 13.

MARC-a nastalih u Sjedinjenim Američkim Državama i Kanadi je nastao format MARC-a danas poznat kao MARC 21. A do kraja 1970-tih je razvijeno više od 20 različitih MARC formata, a 1977. godine IFLA objavljuje UNIMARC (Universal MARC format).[18] "Prvo izdanje UNIMARC-a IFLA objavljuje 1977. godine. Sam naziv formata govori da ga je međunarodna zajednica smatrala univerzalnim formatom za razmjenu bibliografskih podataka na strojno čitljivom mediju."[19]

Kataložnim se pravilima razlikovalo u katalogizaciji odrednica (podaci o autoru ili naslovu publikacije prema kojima se određuje njezino mjesto u abecednom katalogu) od kataložnoga opisa (ostali podaci bitni za identifikaciju određene publikacije).

2.2 Bibliografski zapis i organizacija informacija

Za proces katalogizacije bitni su, dakle, ISBD standard i MARC format. Njima je omogućen standardizirani i razmjenjiv zapis u interoperabilnim sustavima. Temelj postojanja ISBD-a je usklađivanje kataložnih opisa u svrhu olakšane razmjene zapisa među nacionalnim knjižnicama svijeta. „Cilj ISBD-a je osigurati standard kojim bi se mogla provoditi univerzalna bibliografska kontrola, osigurati univerzalnost i trenutnu dostupnost u međunarodno prihvatljivom formatu te pruža temeljne bibliografske podatke za svu vrstu građe u svim zemljama."[20]

Promjene koje se događaju ISBD standardima izravno utječu na katalogizaciju pa time i na bibliografske zapise. U „*ISBD: međunarodni standardni bibliografski opis*" najbolje je opisano što ISBD određuje: "uvjete opisa i identifikacije objavljene građe koja se može pronaći u knjižničnim zbirkama. Standard utvrđuje elemente podataka koji se bilježe ili preuzimaju u unaprijed određenom slijedu kao osnovu opisa građe te rabi propisanu interpunkciju kao sredstvo prepoznavanja i prikaza elemenata podataka što ih čini razumljivim neovisno o jeziku opisa."[21]

Knjižničarska zajednica je težila postići zapise koji su pristupačni, o građi objavljenoj diljem svijeta, svima onima koji te zapise trebaju iz bilo kojeg razloga i s bilo kojeg mjesta. Osim toga, ISBD također pomaže i pri uštedi financijskih resursa knjižnica osiguravajući mogućnost lakše razmjene bibliografskih zapisa. Upravo je razmjena zapisa jedna od pozitivnih

[18] Usp. Willer, Mirna. UNIMARC u teoriji i praksi. Rijeka: Naklada Benja, 1996. Str 16.
[19] Willer, Mirna. UNIMARC u teoriji i praksi. Rijeka: Naklada Benja, 1996. Str. 16.
[20] Isto. Str. XV.
[21] ISBD: međunarodni standardni bibliografski opis / preporučila Skupina za pregled ISBD-a; odobrio Stalni odbor IFLA-ine Sekcije za katalogizaciju; [s engleskog prevela i predgovor napisala Ana Barbarić]. Objedinjeno izd. Zagreb: Hrvatsko knjižničarsko društvo, 2014. Str. 1.

posljedica standardizacije te donosi razmjenu zapisa te interoperabilnost opisa kao važne teme za suvremeno knjižničarstvo.

"Svrha je ISBD-a da utvrđivanjem elemenata od kojih se sastoji bibliografski opis propisivanjem redoslijeda prema kojem se ti elementi moraju prikazivati i, u sladu s tim, propisivanjem interpunkcijskih simbola kojima se elementi moraju odijeliti:

1. omogući razmjena zapisa iz raznih izvora, tako da se zapisi izrađeni u jednoj zemlji lako mogu preuzeti u knjižničarske kataloge ili druge bibliografske popise bilo koje druge zemlje;
2. pomogne u razumijevanju zapisa usprkos jezičnim zaprekama, tako da zapisi izrađeni za govornike jednog jezika budu razumljivi govornicima drugih jezika;
3. olakša prijenos bibliografskih zapisa u elektroničnom obliku;
4. poboljša prenosivost bibliografskih podataka u okruženje semantičkog weba kao interoperabilnost ISBD-a s drugim standardima sadržaja."[22]

ISBD opisi prvenstveno su osmišljeni za kartične kataloge, odnosno informacije koje su korisnicima potrebne kako bi identificirali jedinicu građe i utvrdili njezinu fizičku lokaciju u nekoj zbirci putem kartičnog kataloga. U suvremenijoj sredini ISBD opis zadržava kompaktnost zapisa za potrebe prikaza krajnjem korisniku. Uspoređujući elemente opisa koji se opisuju prema ISBD-u, s Kalimaovim Pinakesima možemo uočiti da se u opisu zapisuju vrlo slične informacije, osim detaljnih informacija o autoru.

Prednost koju je ISBD donio je univerzalnost i izjednačenost zapisa. Korisnici, nakon što je ISBD prihvaćen na međunarodnoj razini, bez obzira u kojoj knjižnici pretražuju građu naići će na iste informacije i isti izgled zapisa. Jedan od nedostataka, usprkos učestalim preradama ISBD-a, je nemogućnost pokrivanja svih bitnih elemenata koje bi neka građa mogla posjedovati i koji bi bili bitni za opis. U tim slučajevima, knjižničar mora bitne podatke zapisati pod napomenama koje nisu standardizirane pa je po njima znatno teže pretraživati. Nažalost, iako ISBD potiče i teži interoperabilnosti bibliografskih podataka možemo uočiti kako izvan tradicionalnih knjižnica dolazi do suprotnog efekta.

Iako razmjena bibliografskih podataka postoji barem od ISBD standardizacije podataka koji se zapisuju na kataložne kartice, razmjena nije bila jednostavna ni jeftina. Bibliografski zapisi su se prenosili ili preko mikrofilmova, ili preko bibliografija ili preko ujedinjenih

[22] Isto. Str. 1.

kataloga ili su preuzimali podatke s CIP (Cataloging in Publication / Katalogizacija u publikaciji) zapisa. Ovakvi katalozi nisu bili jeftini za izraditi, bili su nepraktični za česte promjene i ažuriranja te teški za korištenje.[23] Razvojem računala, Kongresna knjižnica je uvidjela prednost koju bi im dovela automatizacija katalogizacije. "Identificirane bibliografske podatke i njihovu strukturu bilo je potrebno jedinstveno opisati i organizirati u skupove elemenata podataka te odrediti njihove odnose kako bi im se dodijelili standardizirani kodovi u okviru formata za strojno čitljivo katalogiziranje."[24]

Bibliografski zapisi, informacije koje su tradicionalno bile zapisane i dostupne samo na kataložnim karticama, postale su upotrebljive računalu pomoću MARC formata. Upotrebljivost informacija računalu znatno utječe na upravljanje i analizu tih zapisa budući da su računala vrlo dobra u baratanju strukturiranim zapisima na makro razinama. Međutim, postojanje različitih verzija MARC formata škodilo je interoperabilnosti MARC zapisa jer iako su pisani po istoj ili sličnoj sintaksi sami elementi zapisa nisu formirani prema istom standardu što onemogućuje razmjenu.

Jedan od bitnih razvojnih elemenata koji je nastao s razvojem MARC 21 je *Format za razmjenu bibliografskih podataka na magnetskim trakama* (ISO 2709). Ovaj format je omogućio standardizaciju formata zapisa, strukturalnu razmjenu zapisa, konverziju zapisa (interoperabilnost), prijelaz na novu fizičku razinu (što je kasnije postalo bitno za prijelaz u XML format) te je omogućio jasniji pristup u stvaranju baza podataka. Zanimljivo je i primijetiti da je MARC zapis nastao prije ideje relacijskog modela podataka što zasigurno jedan od uzroka poteškoća u pohrani MARC zapisa u relacijske baze podataka za potrebe integriranih knjižničnih sustava.

Format propisuje dijelove koje zapis mora posjedovati za razmjenu: strukturu zapisa, označitelje sadržaja i sadržaj podataka. Podjela oznake polja se dijeli na blokove njihova polja, potpolja i u nekim slučajevima indikatore. Drugim riječima osnovni organizacijski model MARC formata je hijerarhijski. Polja imamo od 0 do 9 koja sadrže tri broja, svako polje onda ima potpolja koja su označena slovima a, b, c i tako dalje.[25] Za neka polja postoje unaprijed određeni podaci koji se zapisuju, a to je pogotovo bitno za blokove namijenjene za računalo, kao što je kod bloka za identifikaciju i bloka kodiranih informacija. Osim toga, treba se

[23] Usp. The History and Purpose of MARC. URL: http://lili.org/forlibs/ce/able/course8/03purpose.htm (2018.05.03)
[24] Willer, Mirna. UNIMARC u teoriji i praksi. Rijeka: Naklada Benja, 1996. Str 14
[25] Usp. Isto. Str 65-67.

napomenuti i činjenicu da je MARC nastao s ciljem automatizacije katalogizacije te je ono zapravo preslika informacija prisutnih na kataložnom listiću. MARC zapis je nemoguće napisati bez predznanja o pravilima izrade kataložnih listića odnosno propisanim interpunkcijama za spajanje i razdvajanje individualnih vrijednosti.

MARC zapis sam po sebi osim bibliografskih podataka ima i podatke koje su namijenjene isključivo računalu i podatke koji olakšavaju katalogizatorima katalogizaciju te upravljanje poslovanjem ustanove. Razvoj UNIMARC formata je usko vezan uz ISBD. Sva polja su prilagođena ISBD-u i sve promijene koje su se događala na ISBD-u su imale utjecaj i na UNIMARC te se on mijenjao sukladno tome. UNIMARC u potpunosti podržava strukturu ISBD-a, svi elementi su podržani s odraženim poljima te je nužno upisivanje ISBD interpunkcija time osiguravajući UNIMARC-ovu prikladnost za razmjenu bibliografskih podataka s drugim knjižničarskim sustavima.[26] Neke od ovih značajki nisu podobne za digitalnu sredinu već bi se trebale odrađivati automatski. Posljedice možemo vidjeti prilikom izvoza podataka iz integriranih knjižničnih sustava gdje često kataložna interpunkcija postaje dio, primjerice, naslova prilikom izvoza u druge formate poput Dublin Core zapisa u RDF formatu.

Osim informacija koje propisuje ISBD, u MARC zapisu može sadržavati i popratne dijelove, kao informacije sadržajne obrade, stručne oznake se označavaju s nekim od knjižničarskih klasifikacijskih sustava (Univerzalna decimalna klasifikacija (UDK), Deweyeva decimalna klasifikacija (DDK, Klasifikacija Kongresne knjižnice (LCC)) i predmetnice.

2.2.1 Bibliografski zapis i sadržajna obrada

Sadržajna obrada, iako jednako važna kao i sama katalogizacija za obradu, tek se u moderno doba krenulo raspravljati o potrebi za standardiziranom i univerzalnom klasifikacijskom sustavu. Ova činjenica je posebno zanimljiva kada znamo da se i dalje građa organizira prema sadržajnoj podjeli. Usprkos ideje o univerzalnoj klasifikaciji, knjižnice i dalje koriste različite knjižničarske klasifikacijske sustave.

Bez obzira koji klasifikacijski sustav neka knjižnica koristi, tim oznakama se šire označi tema jedinice građe. Katalogizator sam odlučuje koliko će duboko ići s oznakama, pošto, npr. u Univerzalnoj decimalnoj klasifikaciji je omogućeno odrediti veću ili manju oznaku pošto je podijeljena na glavnu i pomoćnu tablicu. Glavna tablica u UDK se dijeli na 10 skupina, a 4 skupina je slobodna. Pomoćnim tablicama se označava jezik, oblik, mjesto, etničke grupe,

[26] Usp. Isto. Str. 64.

vrijeme. Iako klasifikacije dopuštaju da katalogizatori samostalno određuju kojoj skupini pripada neka jedinica građe, skupine su dogovorene. Kao što je već navedeno postoji više knjižničarskih klasifikacijskih sustava, u tom smislu imamo sličnu situaciju kao i kod MARC-a. Razlika je utoliko jer korisnik može lakše uočiti koju podjelu koristi knjižnica jer je jasnije naznačeno, za razliku od toga koji MARC koristi. Osim toga, jednostavnije je i pronaći skupine na koje je podijeljena klasifikacijska shema i što znači ta oznaka, nasuprot skupina MARC-a.

Nasuprot klasifikacijskih sustava, predmetnice nisu propisane, već su pružene smjernice o kako se određuju i kako ih se treba zapisivati. Iako nisu striktno propisane preporuča se korištenje nekih od tezaurusa. Često se tijekom katalogizacije, dio sadržajne obrade ili preskoči ili samo preuzme bez dodatnih analiza.

Problematika sadržajne obrade kao i uključivanja sadržajnih oznaka u bibliografski opis je vrlo široka i izvan opsega ovog rada. Ovdje se spominje kako bi se držalo na umu koje su sve informacije važne za razmjenu i interoperabilnost.

2.3 Strukturirani zapis bibliografskih informacija

Bibliografski opisi do nedavno nisu imali posve standardne komponente i ne mogu ih posve ni imati zbog specijalizacije. Ipak, postoje određeni dijelovi koji se nalaze u velikoj većini bibliografskih opisa. To su podaci o naslovima, autorima i ostalim podacima o odgovornosti te vremenu objave. Kroz vrijeme nastaje potreba za standardizacijom ovih zapisa te se ona osigurava dogovorom ISBD-a.

ISBD svoj opis temelji na 9 skupina, gdje svaka osim 7. (Napomene[27]) ima podelemente. Iako je opis strukturiran i detaljno unaprijed određen, osigurana je određena sloboda pri pisanju bibliografskog opisa. Katalogizator ne treba nužno koristiti sve skupine tijekom opisivanja jedinice građe, kao što je u slučaju 3. skupine (Podacima specifičnim za vrstu građe[28]) koja je namijenjena građi kao karte, CD, muzikalije i slično, ili pak 4. skupina (Izdavanje, proizvodnja, raspačavanje[29]) koja se koristi u slučaju pisanja bibliografije. Ne opisuje se idealni primjerak kao u slučaju bibliografskog opisa za bibliografiju, već je ova vrsta slobode omogućena kako bi se što preciznije opisala jedinca građe koju knjižnica posjeduje.

[27] ISBD: međunarodni standardni bibliografski opis / preporučila Skupina za pregled ISBD-a; odobrio Stalni odbor IFLA-ine Sekcije za katalogizaciju; [s engleskog prevela i predgovor napisala Ana Barbarić]. Objedinjeno izd. Zagreb: Hrvatsko knjižničarsko društvo, 2014.
[28] Isto.
[29] Isto.

MARC formati su usko vezani uz ISBD, sva polja su prilagođena ISBD-u i sve promijene koje su se događala na ISBD-u su imale utjecaj na polja koja se trebaju zapisivati te se on mijenjao sukladno tome. MARC u potpunosti podržava strukturu ISBD-a, svi elementi su podržani s odraženim poljima. U MARC zapisima se nalaze isti podaci koje bi zapisali sukladno ISBD-u, s dodanim poljima koji se odnose na potrebe umrežavanja podataka u online katalozima.

Podaci prisutni u bibliografskim podacima su proizvod eliminacije koje su informacije bitne kako bi se što lakše identificirao primjerak. Struktura i informacije su bile podložne vremenu i novim vrstama građe kako su nastajale. Struktura se prilagođava informacijama koje se zapisuju, ali od dogovorenih standarda ISBD-a i formata MARC, struktura je standardizirana i napisana prema predlošku kartičnih kataloga odnosno za MARC format se razvija posebna struktura koja pogoduje računalu i čuva podatke potrebne za izradu kartičnih kataloga. Postoji razlika u izgledu strukture za podatke koji se prikazuju korisnicima od onih koje struka koristi.

Iz ISBD zapisa tipično možemo iščitati naslov knjige, ime i prezime autora, kada, gdje i koja nakladnička kuća je izdala tu knjigu, materijalni opis knjige te dodatne informacije ako su važne za tu jedinicu građe. Ove informacije, iako zadovoljavaju potrebe pronalaska građe, u nekim specifičnim slučajevima, kao u slučaju kada korisnik želi okupiti podatke o građi koju će koristiti, određeni podaci (broj stranica, visina hrpta, napomene i slično) su korisniku nepotrebni u krajnjem prikazu iako ih glavni zapis mora sadržavati. Ovaj problem se kasnije samo prebacio s analognih sustava u digitalne. Osjetljivost na specifične podatke koje bi korisnik mogao koristiti u pretraživanju i koje bi smatrao nužnima možemo pronaći kod opisa posebne građe kao na primjer kod stare građe.

Svaki podatak je potkrijepljen sa svojim elementom u ISBD standardu, a time i u MARC formatu. Neki podaci imaju dodatna polja u MARC zapisu kako bi se omogućilo umrežavanje podataka unutar kataloga.

2.4 Izvori bibliografskih podataka: katalozi i baze

Kroz stoljeća se pojavljuju različite etape izvora bibliografskih podataka. U početku su to bile naslovnice na kojima su nakladnici ispisivali cijelu svoju produkciju, a kasnije su knjižnice imale svoje popise građe. Knjižnični katalozi su prolazili kroz različite promijene sukladno s vremenom, a danas već imamo različite generacije OPAC kataloga. OPAC je online knjižnični katalog, koji se često kritizira kao teško iskoristiv od strane korisnika i često ne daje lako iskoristive oblike izvoza podataka jer knjižničari nisu razvili vlastite ovakve formate koji

pogoduju mrežnom okruženju. Sve popularnije su komercijalne baze podataka koje svoja sučelja prilagođavaju suvremenim trendovima i potrebama korisnika.

Pristup bibliografskim podacima danas je vrlo jednostavan i ubrzan. Komercijalne baze podataka su više manje preuzele ulogu izvora bibliografskih podataka. "Baza podataka je skup međusobno povezanih podataka, pohranjenih u vanjskoj memoriji računala. Podaci su istovremeno dostupni raznim korisnicima i aplikacijskim programima. Ovo je jedan od načina za strukturiranje, pohranu i brzo pristupanje golemoj količini podataka elektroničkim putem."[30] Komercijalne baze znatno više razmišljaju o iskoristivosti (to im je izvor financiranja), ali njima možemo zamjeriti što često pre-pojednostavljuju organizaciju i pristup informacijama te time mijenjaju navike korisnika.

Komercijalne baze možemo podijeliti u: bibliografske baze podataka, baze cjelovitih tekstova i citatne indekse. Sve tri vrste baza se temelje na bibliografskim podacima kao organizacijskoj okosnici. Baze cjelovitih tekstova uz njih vežu i same publikacije, a citatni indeksi ih povezuju prema referencama navedenim u publikacijama. Te vrste baza su zapravo dodatne razine bibliografskih baza.

Bibliografska baza podataka je baza bibliografskih zapisa ujednačenih opisa. To je zbirka digitalnih referenci za knjige, časopise, novine, pravne publikacije, patente i slično. Knjižničarski bibliografski zapisi za određenu građu su često pridruženi nadređenom opisu. Na primjer, individualni članci u načelu nemaju zaseban opis već će biti samo opisan časopis ili novina kao bibliografska jedinica. Nasuprot tome, u online bazama podataka možemo pronaći za svaki rad zaseban zapis s bogatim ključnim riječima, sažetkom i predmetnicama i vrlo često poveznicu za pristup tom istom članku.[31]

Knjižnice i nakladnici su uvijek nalazili načine kako da prikažu podatke o tome što sve posjeduju za potrebe kako korisnika tako i internog upravljanja. U početku su nakladnici svoju nakladu objavljivali na koricama knjiga, a danas imamo sofisticirana sučelja s mogućnosti pretraživanja i automatskog redanja preko bilo kojeg bibliografskog podatka. Knjižnični katalog je skupina bibliografskih opisa danas najčešće zapisanih prema MARC formatu i ISBD standardu. Katalogizacijom nastaju zapisi, a ona kao takva je prožeta već unaprijed postavljenim pravilima i tradicijom. Jedna od tradicija je da se pojedinačni članci ne opisuju,

[30] Bell, Suzanne S. Librarian's Guide to Online Searching: Cultivating Database Skills for Research and Instruction. 4. Izd. Santa Barbara; Denver; Oxford: Libraries Unlimited, 2015. Str 1.
[31] Usp Feather, John; Sturges, Paul, eds. International Encyclopedia of Information and Library Science. 2. izd. London: Routledge. 2003. Str. 127.

već se opisuje časopis odakle potječu. Ovakav način opisa je ostao još od doba kada je bilo skupo izrađivati bibliografski opis.[32] Upravo je mogućnost razmjene interoperabilnih zapisa jedno od važnih rješenja ovog problema.

Online katalozi su razvijeni prema predlošku kartičnih kataloga. Le Beouf ističe za kartične kataloge, "razvijeni su s paradigmom da mogu poslužiti samo za pronalaženje "poznate stavke" ili za odgovaranje na precizan predmetni upit."[33]

Prvi online katalozi su bili kopija kartičnih kataloga. Iako zapisi isto izgledaju i danas, podaci su s novim OPAC sustavima pristupačniji i lakše pretraživi. Prema Hrvatskoj enciklopediji "OPAC (akronim od engl. *Online Public Access Catalogue*), računalni katalog knjižnice namijenjen javnosti i dostupan s bilo kojeg umreženoga računala. Korisnik pristupa katalogu putem jednostavnoga grafičkog sučelja koje omogućuje komuniciranje bez posrednika."[34] OPAC je omogućio automatsko povezivanje podataka. Skoro svaki dio jednog bibliografskog zapisa je pretraživ i povezan. Informacije o tome što knjižnica posjeduje se na sistematski način organiziraju te se sastoje od bibliografske baze i sofisticiranog sustava pretraživanja sukladno bibliografskim standardima.[35] Sve zajedno, uključujući bazu podataka i razna sučelja za profesionalce, korisnike i aplikacije, često nazivamo Integriranim knjižničnim sustavima ili *Integrated Library System* (ILS). Postoje nekoliko ILS softvera koji se koriste u Hrvatskoj s kojima je omogućena izgradnja skupnog kataloga kao što je KOHA koji koristi Sveučilišna knjižnica Filozofskog fakulteta Sveučilišta u Zagrebu i koji je okupio kataloge ranijih pojedinačnih knjižnica raznih odsjeka tog fakulteta, katalog Nacionalne i sveučilišne knjižnice u Zagrebu koje je izveden u sustavu Aleph (*Automated Library Expandable Program Hebrew*) i CROLIST koji je skupni katalog knjižnica u istoimenom konzorciju.

Korištenje OPAC sustava za prikaz bibliografskih zapisa je donijelo mnoštvo prednosti, omogućilo pristup i pretragu zapisa preko naslova, autora, identifikacijskih oznaka i ostalih komponenata bibliografskog opisa te omogućilo uvid u knjižnice koje posjeduju primjerak. OPAC sučelja liče na tipična sučelja za računalni dohvat podataka te omogućuju korisniku

[32] Usp. Bell, Suzanne S. Librarian's Guide to Online Searching: Cultivating Database Skills for Research and Instruction. 4. Izd. Santa Barbara; Denver; Oxford: Libraries Unlimited, 2015. Str 2-3.
[33] Le Boeuf, Patrick. FRBR: Hype or cure-all? Introduction. // Cataloging & Classification Quarterly 39, 3-4 (2005.). Str. 5.
[34] OPAC. // Hrvatska enciklopedija URL: http://www.enciklopedija.hr/natuknica.aspx?id=45194 (2018-05-15)
[35] Usp. Lewandowski, Dirk. "Using search engine technology to improve library catalogs." Advances in Librarianship. Emerald Group Publishing Limited, 2010. Str. 35-54. URL: https://arxiv.org/pdf/1511.05808.pdf (2018-05-16)

preciznu i lakšu pretragu. Knjižnice često sabiru popise s opisom baza podataka kako bi korisnik mogao odabrati koja je baza podataka prikladna za njegove potrebe.

S druge strane, nakladnički katalozi su jedan od načina da nakladničke kuće oglašavaju svoje proizvode. Isto kao i knjižnični katalozi i nakladnički su dolazili u različitim formatima kroz godine. Danas, kao i sve ostalo, su uglavnom dostupni online. Iako omogućavaju pretraživanje prema više-manje istim elementima, nakladnički su u mnogome lakši jer dopuštaju više pogrešaka, ali često i manju preciznost i slabije mogućnosti. Jedna od ključnih razlika je i u tome što pružaju sažetak odnosno sadržajni prikaz djela pa su značajni za korisnikov izbor iako su promotivnog karaktera.[36]

Pojavom online baza podataka počinjemo razlikovati dvije vrste bibliografskih baza podataka. Prva su enumerativne ili opisne koje su zapravo tradicionalni katalozi u online verziji, druge su sustavi ili zbirke zapisa o aktualnim publikacijama ili člancima te često imamo i mogućnost pristupa digitalnom obliku novih radova u bibliografskoj bazi pomoću hyperlinkova kao što su SCOPUS, Web of Science, EBSCO, JSTOR i slično. Ne smiju se isključiti ni baze koje nisu isključivo usmjerene prema bibliografskim podacima kao što su Google, Google Scholar i Amazon koje su opće puno poznatije i uglavnom su besplatne nasuprot bibliografskih baza. Zanimljivo je da je razvidno da mnogi korisnici bibliografske informacije dohvaćaju upravo kroz ove izvore. Organizacije nevezano za knjižnice su krenule razvijati sofisticiranije sustave za dohvaćanje informacija.[37]

Danas već imamo različite vrte online bibliografskih baza, imamo sa slobodnim pristupom, ali i baze na kojima se plaćaju pretplate. Preko baza je omogućen odaziv bibliografskih opisa za dohvaćene dokumente, često je u sklopu opisa i sažetak rada kao i ključne riječi. Ova vrsta baza se osim za pronalaženje i pretraživanje relevantnih referenci koristi i kako bi se ušlo u trag originalnim dokumentima.[38]

Online baze osim popisa publikaciju objavljuju i pružaju pristup. Uz popis i sažetaka djela na popisu, osigurali su mogućnost uvida u citiranost i poveznice na tekst. Baze koje imaju pristup većem broju izvora su baš zbog toga popularnije i često postoji naknada na mjesečnoj

36 Usp. Mason, Moyak. Publishers Catalogues and Collection Development. URL: http://www.moyak.com/papers/collection-development-publishers.html (2018-05-08)

37 Usp. Bell, Suzanne S. Librarian's Guide to Online Searching: Cultivating Database Skills for Research and Instruction. 4.izd. Santa Barbara; Denver; Oxford: Libraries Unlimited, 2015.

38 Usp. Kusserow, Arne; Groppe, Sven. Getting Indexed by Bibliographic Databases in the Area of Computer Science. // Open Journal of Web Technologies (OJWT)1, 2 (2014).

ili godišnjoj bazi za omogućen pristup. Primjeri takvih baza su Scopus, Web of Knowledge i slične zbirke publikacija.

U knjižničnom smislu, sučelja za otkrivanje literature (*discovery* sučelja) su sustavi koji prikupljaju i unaprijed indeksiraju različite knjižnične sadržaje iz različitih izvora. Izvori mogu dolaziti iz knjižničnih baza podataka, online kataloga, različitih repozitorija i slično te automatski. Sličnu metodu koristi i Google, indeksira i pruža trenutno rangiran popis, ali umjesto web stranica pruža provjerene izvore za koje je već neka knjižnica platila. Jedni od poznatijih sučelja su EBSCO Discovery Service, OCLC-ov WorldCat Discovery Service i slično.[39]

Ovakve baze korisnicima često pružaju lakše prikupljanje bibliografskih podataka pošto dopuštaju preuzimanje podataka u suvremenim formatima za razmjenu podataka kao što su to XML i JSON formatu, a osim toga i pružaju unaprijed različite vrste citata za građu koju koriste. "Postoje različiti formati i protokoli za razmjenu. Oni se koriste za razmjenu i unošenje metapodataka o publikaciji. Uglavnom već unaprijed imaju zadana polja i stilove citiranja i koriste se za izgradnju osobne baze bibliografskih podataka."[40]

[39] Usp. Bell, Suzanne S. Librarian's Guide to Online Searching: Cultivating Database Skills for Research and Instruction. 4. Izd. Santa Barbara; Denver; Oxford: Libraries Unlimited, 2015. Str. 7

[40] Kusserow, Arne; Groppe, Sven. Getting Indexed by Bibliographic Databases in the Area of Computer Science. // Open Journal of Web Technologies (OJWT)1, 2 (2014.).

3 Web i faze u razvoju weba

Pored knjižničnog svijeta, u devedesetima se počinje razvijati posve novi informacijski prostor čija pojava se često izjednačuje u važnosti s izumom tiskarskog stroja. "Web ili World Wide Web (www) je informacijski prostor gdje dokumenti i drugi izvori su označeni *Uniform Resource Locators* (URL), međusobno povezani s hipertekst linkovima i dostupni putem interneta."[41] Tim Berners-Lee je dizajnirao web u svrhu olakšane razmjene podataka i dokumenata koji su se nalazili na različitim računalima znanstvenika CERN-a. Drugim riječima izum weba je poduprijet potrebom za drugačijom organizacijom dokumenata, a s vremenom taj izum mijenja kako uopće pristupamo konceptu dokumenta. "Do 1990. godine Berners-Lee je napisao tri ključna elementa koje su i do danas ostale temelji weba:

1. HTML (*HyperText Markup Language*) označiteljski jezik
2. URI (*Uniform Resource Identifier*) jedinstvene oznake za identifikaciju izvora na webu
3. HTTP (*Hypertext Transfer Protocol*) omogućuje dohvaćanje povezanih izvora s cijelog weba"[42]

Iako se internet i web koriste kao sinonimi, to nije slučaj u stvarnosti. Internet je globalan sistem od umreženih računala. Suprotno tome, World Wide Web je nastao kao globalna zbirka dokumenata s vizualnim izričajem koji su međusobno povezani poveznicama (eng. *hyperilink*). Budući da je upravo web uvelike formirao suvremen pristup strukturiranim podacima pa tako i organizaciji informacija (u smislu širem od bibliografske), razvoj weba je za ovaj rad gotovo jednako važan kao i razvoj knjižničnog bibliografskog zapisa.

Kroz različite faze web je imao drugačiji odnos s dokumentima, podacima, ali i s korisnicima tih podataka. Da bi se razumjela važnost interoperabilnosti i razmjene bibliografskih podataka te kako se to može postići, bitno je razumjeti kroz koji je proces prošao današnji najveći izvor podataka kako bi došao do mogućnosti izrazite globalne interoperabilnosti i razmjene podataka. Za potrebe ovog rada možemo razlikovati četiri faze razvoja: klasični, semantički, društveni i web podataka. U literaturi se za svaku fazu može pronaći različite nazive te to dovodi do građe koja opisuje isto, ali pod različitim nazivom.

[41] Architecture of the World Wide Web, Volume One. // W3C Recommendation. 2004. URL: https://www.w3.org/TR/webarch/ (2018-05-08)
[42] History of the Web. // World Wide Web foundation. URL: https://webfoundation.org/about/vision/history-of-the-web/ (2018-05-08)

Klasični web je mreža hipertekstualnih dokumenata. Semantički web je rana ideja označivanja i povezivanja na razini individualnih podataka radije no dokumenata. Umjesto semantičkog weba se u praksi razvija tzv. društveni web: web koji postaje platforma za interaktivne aplikacije, povezuje korisnike i prelazi u „čitaj-piši“ medij. Naposlijetku možemo definirati web podataka, web u kojem ljudi i aplikacije prikupljaju i razmjenjuju podatke na temelju kojih nastaju nove mogućnosti i usluge. Zanimljivo je primijetiti da faza u kojem o webu razmišljamo kao o mreži podataka nije nikad dosegla glavne ideje semantičkog weba koje su formirane još 2001. godine. Osim opisanih viđenja weba zgodno je još spomenuti tzv. web stvari (eng. *web of things*) koji se referira na mogućnost direktnog povezivanja online usluga s objektima iz stvarnog svijeta.

3.1 Klasični web

Nakon što je Berners-Lee 1989. godine predložio otvorenu mrežu računala, prvu uspješnu komunikaciju među web pretraživačima je postignuo već sljedeće godine. Do 1999. godine postoji pozamašan broj web stranica. Stranice su bile statične i samo za čitanje. Mogućnosti koje je pružao Klasični web ili web dokumenata su bile i za očekivati, pošto je web u početku kopirao tisak, a svijest o interaktivnim mogućnostima kao i tehnologija za provedbu istih je tek u razvoju.

Ova faza weba prestaje s propadanjem tzv. *dotcomova,* tijekom 1999.-2001. i posebno s pojavom društvenog weba (tzv. *web 2.0)* oko 2005. godine. Pruža malu mogućnost za interakciju gdje ljudi mogu preuzimati podatke, ali nije bila moguća interakcija s web stranicom. Interakcija i razmjena podataka su u ovoj fazi tek u začetku. Web je tada imao pasivnu ulogu.[43]

Drugim riječima, tada je web služio kao još jedna platforma gdje su tvrtke objavljivale svoje kataloge i brošure kako bi predstavile svoje proizvode, a ljudi su ih mogli pročitati. Iako su stranice mogle pružiti korisne informacije nije postojao razlog da se posjetitelji često vraćaju na njih. Glavni cilj stranica je bio objaviti informacije, korisnici nisu mogli nikako utjecati ili pružiti doprinos, a poveznice se koriste samo za povezivanje dokumenata. Ključni tehnologije su HTTP, HTML i URI kao osnova weba, a ostale važne tehnologije se tek razvijaju (e.g. CSS,

[43] Choudhury, Nupur. World wide web and its journey from web 1.0 to web 4.0. // Internationaln Journal of Computer Science and Engineering 5, 6(2014.). Str. 8096. URL: http://ijcsit.com/docs/Volume%205/vol5issue06/ijcsit20140506265.pdf (2018-05-10)

PHP, JavaScript te promjene paradigme u izradi samih web stranica). Osim toga stranice su bile čitljive samo ljudima, nije postojao sadržaj kompatibilan sa strojem.[44]

Krajnji korisnici su bili jedini koji su imali mogućnost odlučiti da li dokumenti imaju vrijednost. Zapravo, programi koji su podržavali web su manipulirali podacima pohranjenim u kompleksnim strukturama podataka te su pružali pristup samo dokumentima koji imaju određene podatke. Povećanjem broja web stranica došlo je do potrebe za organizacijom kako bi se poboljšao proces pretraživanja. Bez obzira, problem je ostao, krajnjim korisnicima su i dalje bili pruženi dokumenti koji su zadovoljavali upit od par riječi. Tražilice i dalje nisu bile sposobne semantički protumačiti podatke koji su se nalazili u samim HTML dokumentima.

3.2 Semantički web i umreženi podaci

Zanimljivo je da je ideja semantičkog weba nastala vrlo rano u razvoju weba iako je tzv. društveni web u praksi zaživio ranije. Nakon razvoja XML-a kao jezika za označavanje podataka (a za razliku od HTML-a koji služi strukturiranju sadržaja za prikaz) razvija se ideja o povezivanju informacija, a ne samo dokumenata. Ideja Tim Berners-Lee-a kada je osmišljavao semantički web je bila da u web preslika način na koji ljudi razmišljaju i povezuju podatke. U tom slučaju web bi morao razumjeti kontekst podataka kako bi im mogao uspješno povezati. Ključna ideja ovdje je povezivanje na razini podataka, a ne samo na razini dokumenata.

Koncept Semantičkog weba su najbolje objasnili Tim Berners-Lee, James Hendler i Ora Lassila u članku „*The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*" kao "okvir koji omogućava različitim aplikacijama i programima da dijele svoje podatke. Opisan je kao produžetak WWW-a, ali nema istu svrhu kao web. Iako oboje imaju cilj pružanja informacija korisnicima, web pruža korisnicima sve moguće dokumente koji bi bili prikladni njihovoj informacijskoj potrebi dok semantički web klasificira individualne dijelove do razine podataka na način da osigura i računalu i ljudima da koriste podatke na različite načine. Cilj Semantičkog weba je osigurati pristup što većem broju podataka koristeći se sličnim tehnikama kao i arhitektura weba, kao što su URI-i i URL."[45]

[44] Usp. Aghaei, Sareh, Mohammad Ali Nematbakhsh, and Hadi Khosravi Farsani. Evolution of the world wide web: From web 1.0 to web 4.0. International Journal of Web & Semantic Technology 3, 1(2012.). Str 2-3. URL: http://www.ftsm.ukm.my/ss/Book/EVOLUTION%20OF%20WWW.pdf (2018-05-10)

[45] Berners-Lee, Tim. Hendler, James. Lassila, Ora. The Semantic Web. // Scientific American. Scientific American, a Division of Nature America, 2001. URL: https://www-

Tim Berners Lee Semantički web objašnjava kao "produžetak trenutnog weba u kojem se podacima pružaju dobro definirana značenja, kako bi omogućila bolja suradnja između ljudi i računala."[46] Semantički web je bila ideja idealnog weba gdje bi sav njegov sadržaj dijelio globalnu semantiku. Dvije ključne stvari su bili pokretači za razvoj Semantičkog weba. Prvo je činjenica o razmjeni podataka i informacija. Ako se podaci i informacije razmjenjuju na meta-ljudskim razinama jedino su strojevi sposobni za pristup, organizaciji i upravljanje takvim podacima. Drugo je problem dodjeljivanja značenja informacijama.

Osmišljena su dva cilja kako bi se savladala ova dva problema. Prvo su se trebali razviti modeli za opisivanje podataka, shema i relacija koji bi bili dovoljno fleksibilni, računalu čitljiv i jednostavni za distribuciju (RDF i OWL), a drugo je razvoj infrastrukture za web s protokolima, jezicima za postavljanje upita (eng. *query languages*), specifikacije i aplikacije za pristup, savjetovanje, objavljivanje i razmjenu podataka. Kasnije su sve ovo postali ključni elementi za izgradnju tzv. weba podataka.[47]

Ono što se želi postići semantičkim webom zapravo i nije nova ideja. "Teži se tome da računalo/stroj podacima, osim da ih prikazuje na ljudima čitljiv način, pruži tim istim podacima i značenje. Tradicionalni sustavi za reprezentaciju znanja su bili centralizirani, što znači da bi se moglo dijeli znanje svi entiteti moraju imati isto značenje nekog koncepta."[48] Da bi entiteti mogli kvalitetno komunicirati oni moraju dijeliti istu definiciju nekog koncepta, svi moraju znati što se misli pod neki specifični koncept, svi moraju pričat istim jezikom i biti u mogućnosti sudjelovati u kontekstu u kojem se koncept koristi. Da bi se ovo postiglo, trebaju se osmisliti pravila. Kako bi se izbjegli nesporazumi, jezik koji se koristi, osim što treba prikazati podatke, treba i prikazati pravila prema kojima su ti podaci interpretirani i obrazloženi.

Dokumenti na webu su trenutno zapisani u HTML označiteljskom jeziku, ali postoje određena ograničenja. Oznake za metapodatke omogućavaju metodu prema kojoj računala

sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf (2018-05-11)

[45] Isto. Str. 1.

[46] Berners-Lee, Tim. Hendler, James. Lassila, Ora. The Semantic Web. // Scientific American. Scientific American, a Division of Nature America, 2001. Str. 1. URL: https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf (2018-05-11)

[47] Berners-Lee, Tim. Hendler, James. Lassila, Ora. The Semantic Web. // Scientific American. Scientific American, a Division of Nature America, 2001. Str. 1. URL: https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf (2018-05-11)

[48] Kuck, G. Tim Berners-Lee's Semantic Web. // SA Journal of Information Management 6, 1(2004.). Str. 3. URL: https://sajim.co.za/index.php/sajim/article/download/297/288 (2018-05-11)

mogu kategorizirati podatke s web stranica. Njima se mogu prikazati podaci koji opisuju dokument (kao u tradicionalnoj katalogizaciji), ali ne možemo prikazat povezanost na razini informacija koje su zapisane u sam dokument. Sa semantičkim webom dokumenti ne bi bili samo povezani, već bi bili svjesni informacija i njihovih vrijednosti u tim dokumentima.

Arhitektura Semantičkog weba se temelji na određenim elementima koji već jesu sastavni dio weba, ali koji se bi koristili na novim razinama. URI-i su standardi za identifikaciju, URI-i (*Uniform Resource Identifier*) i njegova podvrsta, URL-ovi (*Uniform Resource Locator*), su dogovoreni univerzalni identifikatori koji se koriste kako bi se identificirale web stranice. Njima se identificira gdje se nalazi web izvor u mreži računala i sadrži mehanizam da je dohvati.

Kako bismo ove tehnologije iskoristili za povezivanje podataka (umjesto samo dokumenata) potreban nam je model za opis podataka. *Resource Description Framework* (RDF) je okvir za reprezentaciju podataka na webu. To je jednostavan podatkovni model koji koristi URI-je za identifikaciju izvora i opisivanje odnosa između termina u izvorima u smislu imenovanih svojstva i vrijednosti. RDF pruža generalnu, fleksibilnu metodu za razdvajanje bilo kojeg znanja, podataka u male dijelove, koje zovemo tripleti, s nekim pravilima o semantici (značenju) tih dijelova. Tu je ključna i RDF Shema koja pruža rječnik za modeliranje podataka za RDF podatke. Shema je proširenje RDF rječnika u nekoj domeni znanja. Sljedeće je ključan jezik za označavanja podataka, ili *EXtensible Markup Language* (XML) koji pruža sintaksu prema kojoj se najčešće zapisuju podaci ususatvljeni kroz RDF model. XML je dizajniran kako bi se dokumenti mogli na jednostavan način pohranjivati i odašiljati preko weba i dopušta razvojnom programeru da dodaje značenje podacima koje pohranjuje ili prenosi.[49] Vrijedi i spomenuti da je RDF samo podatkovni model, a XML samo jedan od formata u koji ga možemo zapisati.

Uz samo označavanje podataka, podatke prema tim odnosima moramo moći postaviti u međusobne odnose. Ontologije su način na koji se formalno definiraju odnosi među terminima. Inače se sastoji i od taksonomija i od pravila. Koristi se pošto svaka baza podataka može koristiti drugačiji jezik, druge oznake za identifikaciju nečega u dokumentu. Taksonomije definiraju objekte i odnose među njima.[50] Logika i dokazi su na vrhu ontološke strukture kako bi nove zaključke automatski odvijali, to znači da agenti mogu odraditi dedukciju da li neki

[49] Usp. Kuck, G. Tim Berners-Lee's Semantic Web. // SA Journal of Information Management 6, 1(2004.). Str. 4. URL: https://sajim.co.za/index.php/sajim/article/download/297/288 (2018-05-12)
[50] Usp. Kuck, G. Tim Berners-Lee's Semantic Web. // SA Journal of Information Management 6, 1(2004.). Str. 6. URL: https://sajim.co.za/index.php/sajim/article/download/297/288 (2018-05-12)

izvor zadovoljava njihove potrebe korištenjem različitih sistema za analizu i rasuđivanje. Zadnji sloj za semantički web je povjerenje kojim bi se trebala osigurati kvaliteta informacija na webu i stupanj povjerenja u izvor koji pruža te informacije.[51]

Glavni cilj semantičkog weba je povezivanje podataka, a ne objavljivanje statičnih dokumenata. Godine 2007. Tim Berners Lee predstavlja skupinu pravila koji kasnije postaju poznati kao *Linked Data principles* za objavljivanje i povezivanje podataka na webu. Stvaratelji podataka mogu tako dodati svoje podatke u globalni prostor podataka tako da izdaju podatke na webu sukladno principima *Linked Data* (umreženih podataka).[52]

Linked data ili Umreženi podaci su srce Semantičkog weba, to je tehnika koja omogućava stvaranje Semantičkog weba. Pomoću umreženih podataka kada imamo jedan od podataka možemo pronaći i ostale podatke koji su povezani/srodni. Umreženi podaci su način na koji se koristi web kako bi se stvorili veze između podataka iz različitih izvora. Dok se Klasični web sastojao od HTML dokumenata povezanih s hipervezama, web s umreženim podacima se oslanja na dokumente kojim su podaci zapisani prema RDF modelu. Umreženi podaci trebali bi biti objavljeni zajedno s nekoliko vrsta metapodataka, kako bi se povećala vrijednost za korisnike. Kako bi klijentima omogućili procjenu kvalitete objavljenih podataka i omogućili izbor, podaci trebaju biti popraćeni meta-informacijama o svom kreatoru, njegovom datumu izrade, kao i načinu izrade.[53] Ključ umreženih podataka leži u podacima zapisanim u mrežnom modelu (odnosno, matematički rečeno, modelu grafa) s globalno jednoznačnim identifikatorima koji se mogu dereferencirati, odnosno URI-ima.

Zapravo jedino što se u praksi ostvarilo i zadržalo od ideje Semantičkog weba su Umreženi podaci, pošto je formalan opis značenja bio teško izvediv i to posebno u kontekstu univerzalne standardizacije među narodima, tradicijama i kulturama. Umreženi podaci upotrebljavaju RDF za izradu poveznica bilo kojih stvari u svijetu koji rezultira s pojavom koju možemo nazvati web podataka. Semantički web bi bio na neki način globalna baza podataka i često ga uspoređuju s webom podataka.

[51] Usp. Aghaei, Sareh, Mohammad Ali Nematbakhsh, and Hadi Khosravi Farsani. Evolution of the world wide web: From web 1.0 to web 4.0. // International Journal of Web & Semantic Technology 3, 1(2012.). Str. 7. URL: http://www.ftsm.ukm.my/ss/Book/EVOLUTION%20OF%20WWW.pdf (2018-05-12)
[52] Usp. Isto.
[53] Usp. Bizer, Christian, Tom Heath, and Tim Berners-Lee. Linked data-the story so far. // International journal on semantic web and information systems 5, 3(2009.). Str. 2-6.

3.3 Društveni web

Društveni web (eng. *Social web*) je, u praksi, druga generacija weba. Klasični web je bio statičan i korisnici nisu imali utjecaj na stranice, na Društvenom webu ili web 2.0 stranice su interaktivne i konstantno se mijenjaju i glavni stvaratelj konteksta na webu je korisnik. Ovakav web omogućava interakciju između korisnika preko interaktivne tehnologije, još ga se opisuje kao čitaj-piši web. U ovom kontekstu, dokument postaje aplikacija, a web platforma.

Tim O'Reilly definira web 2.0 kao revoluciju u računalnoj industriji uzrokovanu shvaćanja weba kao platformu i pokušaj razumijevanja pravila za uspjeh na toj novoj platformi. Društveni web je još poznat kao web mudrosti, web usmjeren na ljude, participativni web. S čitanjem i pisanjem web je postao dvosmjeran. Web 2.0 je web kao platforma u kojoj korisnici nemaju istu kontrolu koju su imali na klasičnom, tj. korisnici weba 2.0 imaju više interakcije s manje kontrole.[54] Značajno kod web 2.0 ili Društvenog weba je rast i danas već prevlast društvenih mreža na webu iako s vremenom dolazi do puno veće centralizacije usluga nego u ranim danim društvenog weba.

Ključne karakteristike Društvenog weba su: bogato korisničko iskustvo, sudjelovanje korisnika, dinamičan kontekst, metapodaci, web standard i skalabilnost.[55] Bogato korisničko iskustvo je termin koji se odnosi na ponašanje web aplikacija, korisnici imaju s ovakvim aplikacijama sličan doživljaj kao i kod desktop aplikacija. Kada bi na stranicama u Klasičnom webu korisnici mijenjali išta, kako bi se promijene prikazale to bi iziskivalo da se nova stranica generira. Ovaj problem se riješio s Ajax tehnologijom. Asinkroni JavaScript i XML (AJAX) je omogućio da za neke radnje nije potrebno da se podaci prebacuju na server već se mogu odvijati direktno i samo na korisnikovom računalu[56] te da se nakon promjene ažurira samo dio, a ne cijela stranica.

Korisničko sadržajno označavanje danas je popularno poznato pod „tagiranje" i skoro svako objavljivanje na webu je potkrijepljeno korisnikovim označavanjem informacija. Web standarde World Wide Web Consortium (W3C) kontinuirano razvija. Web stranice koje slijede

---

[54] O'Reilly, Tim. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. // International Journal of Digital Economics 65, (2007.). Str. 17-19. URL: https://mpra.ub.uni-muenchen.de/4578/1/mpra_paper_4578.pdf (2018-05-16)

[55] Usp. Best, David. Web 2.0: Next big thing or next big internet bubble. Technical report, Technische Universiteit Eindhoven, 2006. Str. 2-4.

[56] Usp. Adams, Cameron. AJAX: Usable Interactivity with Remote Scripting. 2005. URL: https://www.sitepoint.com/remote-scripting-ajax/ (2018-05-16)

standarde ili preporuke W3C-a su valjanje, ali istraživanja ukazuje na činjenicu kako nekolicina web 2.0 stranica je slijedilo te preporuke. AJAX tehnologije koriste otvoreni format XML-a i Objektni model dokumenta (DOM) koji su oboje W3C preporuke i time se razlikuju od zatvorenog Flash formata. Skalabilnosti označavaju peer-to-peer aplikacije kao Amazon, Ebay. Ovo su primjeri aplikacija, tj. ideja koje su uspješno iskoristile utjecaj mreže.[57]

Glavne tehnologije i usluge web 2.0 uključuje blogove, stvarno jednostavno pobiranje/prikupljanje (RSS), wikije, oznake, folksonomiju, oznake oblake (*tag clouds*) i društvene mreže. Različiti razvojni alati postoje za izradu blogova, wikija i društvenih mreža. Različiti alati su omogućili jednostavnije, brže i jeftinije usvajanje Društvenog weba. [58]

Asinkroni JavaScript i XML, kao što je već objašnjeno je donio mogućnost da se određeni procesi mogu odvijati na korisnikovom računalu i da se može mijenjati samo dio stranice radije no nanovo usnimavati cijela. AJAX je postignuo ovu promjenu uvođenjem web preglednika. Tehnologija je kombinacija različitih tehnika od kojih je jedna JavaScript. Veliki dio koda JavaScripta bude učitan prije nego što krenemo koristiti web stranicu. Samo mali broj podataka je potrebno da se ponovno učita, a to se postiže asinkronom HTTP zahtjevima. Iako AJAX nije nova tehnologija, ono uspješno spaja stare. Stvarno jednostavno pobiranje/prikupljanje ili *Really Simple Syndication* / RDF Site Summary je dobro definirana sintaksa za dokumente na webu koji reflektiraju ažuriranje na web stranici. Sa RSS-om se promijenio način na koji su korisnici obaviješteni o ažuriranju. Metoda se promijenila od toga gdje su korisnici sami morali nailaziti da se ažuriranje/promjena dogodila u model pretplate na sadržaj. I na kraju, bitno je napomenuti da se koristi RDF (*Resource Description Framework*) za opise izvora.[59]

Za pristup web servisima postoje dva protokola, Protokol za jednostavni pristup objektima (*Simple Object Access Protocol* (SOAP)) i Prijenos reprezentativnog stanja (*Representational State Transfer* (REST)). SOAP je komunikacijski protokol koji je isključivo baziran na XML-u i neovisan je o platformi, a za razmjenu informacija koristi HTTP protokol. Neki su smatrali da je SOAP težak za korištenje pa je REST pružio jednostavniju alternativu. Umjesto korištenja XML-a za postavljanje zahtjeva, REST se oslanja na URL.[60] Aplikacijsko

[57] Usp. Best, David. Web 2.0: Next big thing or next big internet bubble. // Technical report, Technische Universiteit Eindhoven, 2006. Str. 4.
[58] Usp. Aghaei, Sareh, Mohammad Ali Nematbakhsh, and Hadi Khosravi Farsani. Evolution of the world wide web: From web 1.0 to web 4.0. // International Journal of Web & Semantic Technology 3, 1(2012.). Str. 3. URL: http://www.ftsm.ukm.my/ss/Book/EVOLUTION%20OF%20WWW.pdf (2018-05-16)
[59] Usp. Best, David. Web 2.0: Next big thing or next big internet bubble. // Technical report, Technische Universiteit Eindhoven, 2006. Str. 5-9.
[60]Usp. Mueller, John. Understanding SOAP and REST Basics and Differences. SmartBear. 2013. URL: https://blog.smartbear.com/apis/understanding-soap-and-rest-basics/ (2018-05-16)

programsko sučelje ili sučelje za programiranje aplikacija (*Application Programming Interface* API) je skup definicija, protokola i alata za izgradnju aplikacijskih softvera, osigurava mogućnost za programere da iskoriste module bez da imaju pristup izvoru koda. API-i koji su bili otvoreni za slobodno korištenje, ili Open API, su omogućili razvoj Društvenog weba i omogućili su okupljanje podataka iz različitih izvora. API omogućava komunikaciju između dvije aplikacije.[61] Vrlo važna novost u tretmanu weba kao platforme je što je korisnicima i aplikacijama omogućena daljinska automatizacija raznih procesa te preuzimanje podataka što podupire dostupnost za potrebe kasnije analize.

U Društvenom webu nasuprot klasičnog, korisnik ima glavnu ulogu u stvaranju sadržaja, ali i vrijednosti. Zbog toga što korisnici pridonose svojim znanjem dobio je različita imena i fraze koje su vezane uz njega, kao kolektivna inteligencija ili mudrost gomile. Time što je omogućeno da korisnici sami stvaraju materijale, može se primijetiti i rast novih zanimanja, kao *blogeri, influenceri, vlogeri...* Osim toga, omogućeno je objedinjenje globalnog znanja; tagiranje slika na raznim društvenim stranicama kao Flickr i Instagram, pisanje članaka za Wikipediju. Ne samo da su informacije objedinjene već je i olakšan i u većini slučajeva besplatan pristup masivnoj količini znanju koje nikad do sada nismo imali.

Ranija ideja Semantičkog weba je ostala prisutna, a društveni web je dao novu nadu mogućnosti kolektivnih umreženih i semantički označenih podataka. Ključ bi bio odnos između ljudi i strojeva. Ljudi bi bili oni koji stvaraju i konzumiraju znanje, a strojevi su oni koji omogućavaju to. Pošto ljudi uče i napreduju kao zajednica komunikacijom, a web bi im trebao osigurati efikasniju komunikaciju. Web osigurava razmjenu znanja koje je spremljeno, pretraživo i moguća je daljnja razmjena.[62] Iako današnji sustavi nisu došli ni blizu razine koja se obećavala u prvom desetljeću 21. stoljeća, web na kojem korisnici i aplikacije razmjenjuju podatke je postao stvarnost. Semantika, međutim, još uvijek izostaje.

3.4 Web podataka

Web postaje globalni izvor znanja koje se koristi za pronalaženje činjenica, rješavanje problema, pronalazak odgovora na pitanja i da se stvori novo znanje. Povezivanjem globalnog znanja omogućena su mnoga nova znanstvena otkrića. Cilj novog weba podataka je podržati globalnu podatkovnu infrastrukturu, u kojoj se stvari iz stvarnog svijeta (entiteti) opisuju na

[61] Usp. Andersen, Per. What is Web 2.0? Ideas, technologies and implications for education. // ISC Technology and Standards Watch 1, 1(2007.). Str. 31.
[62] Usp. Gruber, Tom. Collective knowledge systems: Where the social web meets the semantic web. // Web semantics: science, services and agents on the World Wide Web 6, 1(2008.). Str. 4-5.

webu pomoću strukturiranih podataka, a ne klasičnih dokumenata. Očekuje se da će ova generacija web infrastrukture transformirati način na koji se strukturirani podaci koriste i imati ključnu ulogu u ekonomiji koja se temelji na podacima.[63]

Napretkom poslovnih sustava i računalnih znanosti u posljednjem desetljeću, perspektiva na webu se postupno mijenjala s dokumenata u središtu pozornosti na podatke. Web kao prirodna platforma za upravljanje podacima je dobio glavnu ulogu u novom razvoju. Iako web može pružiti mnogo prednosti, do nedavno principi koji su omogućili webu dokumenata da procvjeta se nisu primijenili i na webu podataka ili umreženih podataka.[64] Tradicionalno su se podaci na webu objavljivali kao sirovi komadi u formatima kao što je CSV (Vrijednosti odvojene zarezom / *Comma-separated values*) ili HTML tablicama žrtvujući time strukturu i semantiku te otežavajući mogućnost dohvaćanja.

Web je nastao s idejom da bude globalni informacijski centar, što je također i jedan od njegovih važnih svojstava, univerzalnost. U početku je bio izazov izgraditi sustav koji bi osigurao univerzalno konzistentno sučelje za strukturirane informacije, dostupne na što više platformi i postepeno razmještene, očekujući da će se novi ljudi i organizacije uključivati. Povezivanjem HTML dokumenta linkovima ili poveznicama imali smo web dokumenata, a danas web masovno funkcionira razmjenom strukturiranih podataka među softverskim agentima, bilo da se radi o pobiranju podataka ili pak o upravljanju web aplikacijama.

Ironično, ono što je donijelo uspjeh webu je također i slabost njegova jezika HTML-a. Nešto slobodnija struktura ovog formata i pristup koji su web preglednici zauzeli prema toleranciji grešaka omogućava da se prikažu i loše napravljene stranice, a HTML je sam po sebi orijentiran prema prikazu sadržaja, ne prema strukturiranju i kontroli. Nova generacija jezika, XML, popravila je oba aspekta. Danas iako znamo da su dokumenti bitan dio globalnih podataka, svjesni smo postojanja mnoštva podataka koji nisu organizirani u obliku dokumenata: tablica podataka, sirovi podaci, slike, podaci sa senzora. Ne smije se zaboraviti i utjecaj Društvenog weba koji je imao na stvaranje golemog broja podataka. Ovi fenomeni su postavili

[63] Usp. Christophides, Vassilis, Vasilis Efthymiou, and Kostas Stefanidis. Entity resolution in the web of data. // Synthesis Lectures on the Semantic Web 5, 3(2015.). Str. 1.
[64] Usp. Bizer, Christian, Tom Heath, and Tim Berners-Lee. Linked data: the story so far. // Semantic services, interoperability and web applications: Emerging concepts. IGI Global, 2011. Str. 1-3.

podatke u centar računalnih disciplina bilo na razini sustava, arhitekture, komunikacije ili razini modeliranja i programiranja.[65]

Sve veći je broj korisnika koji opisuju objekte iz stvarnog svijeta (osobe, mjesta, događaji) kao umrežene podatke. Uspoređujući web dokumenata i web podataka, u webu dokumenata linkovi su među dokumentima, podaci su nestrukturirani, implicira semantiku i predviđen je za isključivo ljudsko korištenje. Nasuprot, web podataka ima poveznice među informacijskim resursima (koji mogu i ne moraju biti dokumenti), podaci su strukturirani te pogodni za računalo, ali čitljivi i ljudima. Jedno od pitanja je bilo kako identificirati ove promjene u upravljanju s podacima kada se uključuju sirovi podaci u model weba dokumenata.

Web podataka je termin koji se često koristi umjesto Semantičkog weba te je teško uočiti granicu između toga što je zapravo web podataka, a što je Semantički web skriven pod nazivom web podataka. Web podataka je zapravo nastao u praksi kao pojednostavljena verzija Semantičkog weba. Često možemo u literaturi pronaći da su baš elementi koji su usko vezani uz ideju Semantičkog weba, kao umreženi podaci, RDF, URI-i, zaslužni za nastanak weba podatka. Web kakav danas imamo je kombinacija ostvarivih i održivih vizija o webu koje je uvelike razvila industrija. Zanimljivo je da je web „otvorio“ podatke putem kojekakvih API sučelja i podaci su dostupni upravo u formatima koji podupiru semantički web, ali su rijetko kada dostupni zapisani prema RDF modelu s pripadajućom shemom i ontologijama.

3.5 Web i podaci

Visoka razina trenutne produkcije, dostupnosti i korištenja podataka upućuje na to da svjedočimo fundamentalnoj promjeni u pristupu informacijama. Tijekom posljednjeg desetljeća može se uočiti značajan preokret u pristupu pojedinaca prema korištenju i razmjeni informacija u formaliziranom obliku razmjenjivih strukturiranih podataka. Dapače, promjene u arhitekturi weba u prelasku društveni medij su prouzrokovale da normalan svakodnevni rad korisnika proizvodi podatke, a rezultirajući skupovi podataka postaju vrlo skup komercijalan proizvod i kompetitivna prednost. Promjena je također vidljiva i u načinu na koji organizacije i tvrtke prikupljaju informacije o svom poslovanju na internetu. “Podatak, poznata ili pretpostavljena činjenica na osnovi koje se oblikuje informacija. Sastoji se od skupa kvantitativnih parametara koji se mogu zapisati kao nizovi znakova ili nizovi brojeva. U

[65] Usp. Gutierrez, Claudio. Modeling the Web of Data (introductory overview). // Reasoning Web. Semantic Technologies for the Web of Data. Springer Berlin Heidelberg, 2011. Str. 5-11. URL: http://citeseerx.ist.psu.edu/viewdoc/download?rep=rep1&type=pdf&doi=10.1.1.221.7285 (2018-05-18)

računalima, koja su do nedavno nazivana i strojevima za automatsku obradbu podataka, ti se nizovi, radi pohrane, obradbe i sl. pretvaraju u nizove bitova."[66]

Za val podataka koji je nastao je zaslužan *World Wide Web* (www). Web bez podataka je teško zamisliti. Web je omogućio radikalnu promjenu u načinu na koji dijelimo podatke, skidajući barijere za objavljivanje i pristup dokumentima kao globalni informacijski sustav. Istraživanja ukazuju na povećanje znanstvenih podataka dvostruko svake godine, pogotovo uspoređujući rezultate s prošlim desetljećem. Neka područja se suočavaju s masovnim povećanjem količine podataka preko satelita, teleskopa, instrumenata visokog kapaciteta, senzorskih mreža, akceleratora i superračunala. Ovaj fenomen nije samo zahvatio znanstvena područja već se može uočiti skoro u svim društvenim sferama, a posebno u poslovanju i upravljanju.

Kroz otprilike dva desetljeća popularnosti weba može se uočiti distinkcija u odnosu na pristup podacima. Kao novi medij, web u početku kopira starije, ali s vremenom razvija vlastiti izričaj i osebujnosti. Web kao mreža dokumenata postaje web kao mreža aplikacija, korisnika i podataka. Danas web stranice korisnici ne doživljavaju kao dokumente, već kao web aplikacije. Web aplikacije više nisu dokumenti već web platforme te je podatke na njima moguće umreživati i povezivati i preživljavaju strukturirani podaci. Na primjeru popularnih društvenih web aplikacija koje omogućuju međusobno povezivanje možemo vidjeti savršen primjer umrežavanja podataka. Podaci u dokumentima postavljeni na webu dokumenata se nisu mogli umreživati niti razmjenjivati, a današnje aplikacije komuniciraju živim podacima. Korisnici u novim fazama weba, web podataka i web korisnika, postaju informacijski stručnjaci. Kada preuzimaju bibliografiju ne žele preuzeti dokument, već želi strukturirane bibliografske podatke pošto imaju potrebu upravljanja njima na strukturiranoj razini radije no na razini običnog računalnog teksta. Činjenica da knjižnice imaju vlastitu porodicu specijaliziranih formata za ovu potrebu (MARC) krajnjim korisnicima ne pomaže puno. U svakom slučaju, u razvojnom webu korisnici laici imaju sve veće potrebe upravljanja informacijama i žele koristiti strukturirane podatke koji su interoperabilni.

[66] Podatak.// Hrvatska enciklopedija. URL: http://enciklopedija.hr/Natuknica.aspx?ID=48887 (2018-05-09)

4 Interoperabilnost i razmjena podataka

Cilj weba je od samih početaka bio omogućiti korisnicima razmjenu dokumenata, a s vremenom se arhitektura weba počinje sve više bazirati na razmjeni podataka. Korisnici s vremenom također sve više počinju funkcionirati na razini razmjene podataka radije no pristupa dokumentima. Sukladno razvoju web tehnologija sve više se informacija objavljivalo na web. Problem je nastao razvojem Društvenog web, trenutak kada je omogućeno da korisnici različitih znanja objavljuju podatke. Imamo pojavu gdje različiti razvojni programeri koristili različite alate, standarde i sheme za zapis i objavu tih podataka. Iako W3C objavljuje preporuke kako zapisivati podatke, to su i dalje samo preporuke na razini formata, a ne upotrebe tih formata u nekoj domeni znanja. Drugim riječima, razvoj shema podataka je ostavljen krajnjim korisnicima, a praksa je češće razviti svoju shemu koja odgovara točnim potrebama radije nego koristiti neku široko dostupnu standardnu shemu. Iz pogleda bibliografskih metapodataka, ne postoji adekvatna i široko prihvaćena ovakva shema.

"Sukladno razvojem različitih informacijskih sustava, svaki novi sustav je razvio svoj informacijski sistem i bazu podataka. Pošto su arhitekture sustava, operativni sustavi i baze podataka bile različite za svaki sistem nisu bili u mogućnosti dijeliti informacije i međusobno komunicirati." Iako određeni sustavi mogu biti u kompatibilni za razmjenu, odnosno komunicirati putem istog formata, to ne treba nužno značiti da su sheme interoperabilne. [67]

Postoje tri razine interoperabilnosti: temeljna ili tehnička, strukturalna ili sintaksna i semantička. Temeljna ili tehnička interoperabilnost omogućava razmjenu podataka između informacijskih sustava, ali ne zahtjeva sposobnost od sustava koji je primio podatke da ih i interpretira. Osigurava osnovne tehničke mogućnosti povezivanja sistema i usluga. Uključuje ključne aspekte kao otvorena sučelja, usluge međusobnog povezivanja, integraciju podataka i softvera, prezentaciju i razmjenu podataka, pristup i sigurnost usluga. Nakon uvođenja Internet protokola (TCP/IP) najosnovnija tehnička interoperabilnost više nije bila problem. Temeljna osigurava samo da se pravilno prenesu podaci, ali nema podataka o tome što se razmijenilo, ne opisuje niti što se razmijenilo (video, slika, podaci) niti pruža smisao podacima. Ovdje se radi o sirovom prijenosu bitova. Za pružanje informacija o prenesenim elementima potrebni su nam

[67] Usp. Zhang, Cui-xiao, et al. Data Exchange based on Web Services. // International Journal of Computer Science and Network Security 6, 5A (2006.). Str. 210.

standardi na sintaksnoj razini.[68] HTML je takva sintaksna i semantička razina, ali isti je nastao samo za potrebe prikaza sadržaja. Za naše potrebe bolji je primjer XML koji je nastao upravo kao standard za opis strukture i značenja, ali on je nastao upravo na uviđenim nedostacima HTML-a.

Kao što vidimo, strukturalna ili sintaksna interoperabilnost je srednja razina koja definira strukturu ili oblik razmijenjenih podataka, uglavnom opisuje format podataka. Podaci koji se prenose komunikacijskim protokolima trebaju imati dobro definiranu sintaksu i kodiranje teksta. Postoji mnoštvo protokola koji prenose podatke, a oni se zapisuju koristeći sintakse poput HTML, XML, a danas je vrlo popularan JSON format. I dalje nam podaci, iako sintaktički definirani i tehnički preneseni, nemaju smisao. Sintakse su u računalstvu gramatika i pravila kako definirati podatke, dok semantika pruža značenje. Samo kada se semantika o podacima definira i prenese, podaci mogu biti procesirani i prepoznati na drugom sistemu. Dok standardi za tehničku i sintaktičku interoperabilnost omogućuju neovisnu razmjenu podataka, semantička interoperabilnost je vrlo specifična za aplikaciju, a time i ovisno o sadržaju specifičnom za uslugu.[69] U ovom smislu je najpotpuniji format XML. HTML ima zatvorenu shemu i jasnu namjenu. JSON i XML služe zapisu strukturiranih formata bilo koje vrste. Samo XML, međutim ima jasno definirano kako se propisuje shema XML zapisa. Kada za dva izvora utvrdimo da koriste istu validacijsku shemu tada su podaci posve interoperabilni.

Drugim riječima, da bi podatke mogli nesmetano razmjenjivati potrebno je da nam oblik u kojem prenosimo podatke zadovoljavaju tri razine interoperabilnosti. Iako je tehnička s Internet protokolom osigurana tako da nesmetano možemo razmijeniti podatke, problemi dolaze kada prijeđemo na sintaksnu i semantičku razinu. Kreatori mogu birati koji će označiteljski jezik koristiti, što se automatski odnosi i na sintaksu koju koristi. Zbog mogućnosti izbora nije sigurno da li će sustav s kojim se razmjenjuju podaci prepoznati sintaksu ili, pak, koristi neku drugu. Sustavi mogu biti interoperabilni, ali to ne znači i uspješnu razmjenu podataka. Kada je u pitanju semantika još uvijek nije u potpunosti osmišljeno kako da se riješi problem da strojevi nauče sami dodjeljivati smisao podacima koji se prenose, tj. da razumiju kontekst podataka i sukladno tome ih povezuju, na čemu se još uvijek ruši ideja posve semantičkog weba.

[68] Usp. Kubicek, Herbert, Ralf Cimander, and Hans Jochen Scholl. Organizational interoperability in e-government: lessons from 77 European good-practice cases. Springer Science & Business Media, 2011. Str. 2-3. URL: https://www.springer.com/cda/content/document/cda_downloaddocument/9783642225017-c7.pdf?SGWID=0-0-45-1229048-p174132406 (2018-05-18)

[69] Usp. Isto. (2018-05-18)

Za uspješnu razmjenu podataka nije bitna samo da su sustavi interoperabilni, jer kao što smo vidjeli sustavi mogu biti inteoperabilni i da se podaci tehnički razmjene. Prava razmjena leži u trenutku kada su podaci preneseni i razmijenjeni na način da će ih osobe moći interpretirati i koristiti. Premda cilj da se ostvari sistem gdje bi kolektivno znanje bilo interoperabilno i lako razmjenjivo ostaje. Da bi razumjeli kako osigurati da podaci budu i uspješno razmijenjeni, a ne samo da su sustavi tehnički interoperabilni, potrebno je razumjeti suvremene formate koji se koriste za zapisivanje i time i razmjenu podataka.

5 Formati za razmjenu podataka

Da bi razmjena podataka bila moguća potrebno je razumijevanje sintakse formata u kojem su podaci zapisani te korišteni model i shema. Bitno je znati kako su podaci zapisani, koja je sintaksa korištena i da li će je naš sustav razumjeti.

Sintaksa računalnih jezika, kao i u lingvistici, je skupina pravila i principa koji određuju strukturu rečenica. Taj skup pravila definira kombinaciju simbola kako bi se dokumenti smatrali pravilno strukturiranim. Računala da bi razumjela što je zapisano očekuju pravilan zapis koji ona „razumiju“ budući da je sintaksa strogo i nedvojbeno definirana. Svaki program definira svoja sintaktička pravila koja kontroliraju što će računalo razumjeti, koje su kombinacije riječi smislene i koje interpunkcijske znakove je potrebno koristiti.[70]

Formati za razmjenu podataka su se razvili kako bi prenosili podatke među sustavima i platformama. Format za razmjenu podataka je neovisan o softveru (bazira se na običnom računalnom tekstu) i domeni znanja i može se koristiti za bilo koju vrstu podataka.

Novi formati za razmjenu podataka nemaju propisanu semantiku i vokabular, odnosno nisu vezani za neku specifičnu domenu. Može se dogoditi da su sustavi interoperabilni na razini modela i sheme opisa, ali razmjena će biti neuspješna ako im je struktura podataka nepoznata. Razmjena podataka je neuspješna kada se podaci ne mogu interpretirati od strane sustava koji ih je preuzeo. Navedeno se dešava radi manjka podrške za neki format ili pak radi različitih modela i vokabulara kojima se potencijalno iste informacije zapisuju u isti format (npr. bibliografske informacije zapisane u XML oblik).

Neki od popularnih formata za zapis i razmjenu podataka su (redom pojavljivanja): CSV, XML i JSON. Većina novih formata za razmjenu je tekstualnog formata kako bi bili neovisni. Podaci odvojeni zarezima (*Comma Separated Values* (CSV)) je poznat i često korišten format zapisa koji služi razmjeni tabličnih podataka. Ovo je najstariji i naizgled jednostavan format, ali sa značajnim nedostacima koji priječe korištenje za kompleksnije podatke poput bibliografskih. Za uspješno iščitavane podataka potrebne su dodatni informacija, tj. objašnjenja koji stupac označava što, inače razumijevanje podataka je nemoguće. Osim toga, CSV struktura je praktična kada zapisujemo dvodimenzionalne podatke jer ne dopušta dodatne raščlambi ili ugnježđivanje podataka.

[70] Usp. Syntax. // Webopedia. URL: https://www.webopedia.com/TERM/S/syntax.html (2018-05-20)

Trenutno preferirani formati za razmjenu podataka su Proširivi jezik za označivanje (*eXtensible Markup Language*, XML) i JavaScript objektna notacija (*JavaScript object notation*, JSON).

5.1 XML

Jedan od formata u kojem podaci mogu biti uspješno razmijenjeni je Proširivi jezik za označivanje ili „*Extensible Markup Language*“ (XML), stvoren tako da je jednostavan za pisanje i čitanje ljudima i računalima.

XML je označiteljski jezik koji se koristi za strukturiranje i opisivanje podataka. Kao vrsta označiteljskog jezika, XML koristi oznake (eng. *tag*) za definiranje elemenata u dokumentima, tj. da pruži informacije o podacima. XML ne ovisi o nekom specifičnom softveru ili hardveru. Zapisuje se kao običan tekst (eng. *plain text*) što ga čini izrazito interoperabilnim. XML je čitljiv i ljudima pošto sadrži i standardne riječi, nasuprot tipičnih programskih sintaksa. Drugačiji je od ostalih označiteljskih jezika jer nema propisan rječnik odnosno shemu već samo propisuje kako strukturirati podatke i stvarati sheme. Stvoren je kako bi čuvao i prenosio podatke.[71] Korisnici XML-a, dakle sami stvaraju sheme (XSD) svojih podataka ili koriste već objavljene standardne sheme.

Goldfarb ističe kako je „XML sveti izvor za kompjuterizaciju, rješavanje problema univerzalnosti razmjene podataka između sličnih sistema.”[72] XML je jednostavan i fleksibilan format proizašao iz SGML označiteljskog jezika iz kojeg je potekao i HTML. Ističe se da sam XML nije označiteljski jezik već metaoznačiteljski jezik pošto on omogućava da se stvore novi označiteljski jezici za specifičnu upotrebu.

XML je originalno dizajniran kako bi omogućio sustavnu razmjenu velikog broja podataka različitih vrsta podataka na webu i drugim sustavima. Uspoređujući HTML s XML-om, HTML dodaje informacije o formatu i tekstu dok XML pruža informacije o podacima koji su zapisani unutar oznake. Cilj HTML-a je da prikaže tekst i multimediju, a XML-u da opiše podatke. Osim toga, još jedna od ključnih razlika je ta što XML ne pruža unaprijed definirane oznake nasuprot HTML-a. XML se može koristiti i kako bi odvojio podatke od drugih

[71] Ups. Markup Language. // The Tech Terms Computer Dictionary. URL: https://techterms.com/definition/markup_language (2018-05-20)
[72] Anderson, Tim. Introducing XML. 2004. URL: http://www.itwriting.com/xmlintro.php (2018-05-20)

elemenata s web stranice, npr. mogu se samo sačuvati XML podaci s web stranice i odvojiti od HTML-a te može ažurirati podatke bez da se mijenja HTML.[73]

Pored pružanja dodatnih mogućnosti strukturiranja web stranica, još jedna od važnih upotreba XML-a je razmjena podataka. U slučajevima kada podatke treba razmijeniti između dva sustava koji nisu kompatibilni, XML može omogućiti razmjenu. XML ne ovisi o specifičnoj vrsti softvera već je kompatibilan s različitim programima. Može pomoći u slučaju kada programi nisu kompatibilni, ali i u slučajevima kada se program ažurira.

XML se također može primijeniti kada označavamo dokument. U tom slučaju treba se zapamtiti da je prva linija koda deklaracijska linija u kojem se definira koja verzija XML-a se koristi i u slučaju kada se koristi oznaka za otvaranje s istom oznakom se treba i zatvoriti. Osim tog, postoje još neka pravila kada je u pitanju XML sintaksa. „Pravila su logična, jednostavna za naučiti i koristiti.

- XML dokument treba sadržavati korijenski ili roditeljski element svih ostalih elemenata,
- XML oznake su osjetljive na velika i mala slova,
- oznaka koja otvara i zatvara element treba biti napisana u istom obliku u oba slučaja,
- treba paziti na redoslijed u slučaju kada koristimo više označitelja za iste podatke, tj. elementi moraju biti pravilno ugniježđeni,
- atributi uvijek moraju biti pod navodnicima i ne smije se zaboraviti da neki simboli imaju svoje uloge u zapisivanju XML-a, kao simbol <.

U slučajevima kada se definirani elementi nalaze u samom tekstu treba se napisati kao referenca kako bi se izbjegla greška. Nasuprot HTML-a ne odstranjuje višak praznina.“[74] Sintaksa XML-a je dakle postrožila sintaksu koju je koristio HTML te se od tada HTML preferira pisati po strožoj, XML sintaksi.

XSD ili XML definicija sheme je preporuka koja specificira kako formalno objasniti elemente u jednom XML dokumentu. Propisuje rječnik, odnosno korištena „polja“ kao i dozvoljene vrijednosti za pojedino polje. XSD služi verifikaciji individualnih XML zapisa. To se može objasniti kao skupina pravila kojim se osigurava da je neki XML dokument validan

[73] Usp. Introduction to XML. // w3schools.com. URL: https://www.w3schools.com/xml/xml_whatis.asp (2018-05-20)
[74] XML Syntax Rules. // w3schools.com. URL: https://www.w3schools.com/xml/xml_syntax.asp (2018-05-20)

prema toj shemi.[75] Drugim riječima, definiraju se obilježja objekata i njihovi međusobni odnosi, a shema predstavlja međusobnu vezu između elemenata, atributa i vrijednosti XML zapisa. Ovime je omogućeno da podaci budu uspješno razmijenjeni pošto je osigurano da oba sistema koja komuniciraju razumiju kako su definirani podaci i da se podatkovni zapisi mogu automatski validirati.

XML je donio mnoštvo prednosti i olakšao je prijenos podataka, ali XML podaci trebaju biti raščlanjeni i analizirani prije nego što budu obrađeni i na strani klijenta i na strani servera. Aplikacije su ukazale na to da taj proces oduzima previše vremena i memorije pošto je potrebno je da se preuzme cijeli XML dokument prije nego se podaci obrade, ali i zbog niske učinkovitosti očitavanja i raščlambe XML podataka tijekom izvršavanja usluge smanjuje se učinkovitost internetskih usluga. Također, XML i vezane tehnologije su relativno kompleksne kada se sve uzme u obzir i u praksi često prekompleksne jer XML ima najviše smisla ukoliko se konzistentno koristi cijela porodica tehnologija (uključujući, na primjer, XML shemu i transformacije). Kada se prenose jednostavni podaci na korisnički zahtjev za potrebu programske upotrebe, pokazalo se da se to može učiniti i formatom direktnije primjerenim za programiranje. Tu se JSON iskazao kao format i sve je popularniji jer je jednostavan za čitanje i pisanje ljudima i jednostavan za procesiranje računalima.[76]

5.2 JSON

JavaScript objektna notacija (eng. *JavaScript Object Notation* (JSON)) je standard za razmjenu podatkovnih objekata. JSON je jednostavan i otvoren tekstualni format koji ima jednostavnu sintaksu što ga čini čitljivi ljudima, ali koristi elemente koji su u sukusu s algoritamskom obradom podataka. Često se koristi u asinkronoj komunikaciji između servera i preglednika. Često ga se uzima kao zamjenu XML-u, iako u nekim slučajevima XML može više (na primjer, označivati tekstove pisane prirodnim jezikom). Ipak, ti slučajevi nisu relevantni za „razmjenu strukturiranih podataka“. Osim što je JSON izrazito čitljiv ljudima, također je jednostavan za analizu računalima. Izgrađen je na dvjema strukturama: „objekt (gdje je objekt skup ključ-vrijednost parova, odnosno vrijednosti se mogu identificirati nazivom) i popis vrijednost (gdje je popis, kao što ime kaže, struktura u kojoj se individualne vrijednosti identificiraju

[75] Usp. XSD (XML Schema Definition). // SearchMicroservices. URL: https://searchmicroservices.techtarget.com/definition/XSD-XML-Schema-Definition (2018-05-20)
[76] Usp. Peng, Dunlu, Lidong Cao, and Wenjie Xu. Using JSON for data exchanging in web service applications. // Journal of Computational Information Systems 7, 16 (2011.). Str. 1. URL: https://www.researchgate.net/publication/265874991_Using_JSON_for_Data_Exchanging_in_Web_Service_Applications (2018-05-20)

redoslijedom).“[77] Ideja za JSON je, dakle, nastala u praksi kao svrsishodan način za komunikaciju među aplikacijama i klijentima. JSON je zapravo jednostavna tekstualna serijalizacija objekata kako se koriste u programskom kôdu.

Što se samih vrijednosti tiče „temeljne vrste podataka su: Booleove vrijednosti, broj, niz znakova (*string*), lista vrijednosti (*arrays*), objekt i null.“[78] Ovakav koncept direktno pruža mogućnost hijerarhijskog modela podataka (MARC je, na primjer, hijerarhijski format) na način koji direktno odgovara konceptima u većini suvremenih programskih jezika kojima se danas obavlja rad upravljanje podacima.

Isto kao i kod XML-a, postoje pravila za sintaksu. „JSON sintaksa se temelji na JavaScript sintaksi:

- podaci su u kombinaciji ime-vrijednost,
- podaci se odvajaju zarezima,
- vitičaste zagrade sadrže objekte, a uglate sadrže listu vrijednosti (arrays),
- njegovi ključevi moraju biti niz znakova i pod navodnicima.“[79]

Uspoređujući ga s XML-om značajno je učinkovitiji u raščlambi podataka. JSON se predstavlja kao alternativa XML-u pošto oba formata imaju podršku za stvaranje, pisanje i dekodiranje u situacijama u stvarnom svijetu gdje se često koriste. XML se koristio za opis strukturiranih podataka i za serijalizaciju objekata i bogatija je porodica standardan. JSON je kraći, jednostavniji za čitanje, ne koristi oznake za zatvaranje linije već koristi koncepte direktno vezane uz suvremene programske jezike. Navedena svojstva i činjenica da je „lagan“ računalima za obradu, su ga postavili u preferiranu poziciju kao format za zapis podataka. Ipak, XML ima šire standarde i povoljniji je u uvjetima u kojima se podaci žele strogo kontrolirati globalnim standardima. U primjerima možemo vidjeti razliku između JSON i XML zapisa.

Primjer 1: JSON

```
{
  "osoba": {
          "ime": "Marko",
          "prezime": "Marulić",
          "OIB": "55554444222",
          "rođendan": {
                  "dan": 23,
                  "mjesec": 4,
                  "godina": 2180
          },
          "mjesto": "Split",
```

[77] Introducing JSON. // Json.org. URL: http://json.org/ (2015-05-20)
[78] JSON Dana Types. // w3schools.com URL: https://www.w3schools.com/js/js_json_datatypes.asp (2018-05-20)
[79] JSON Syntax. // w3schools.com. URL: https://www.w3schools.com/js/js_json_syntax.asp (2018-05-20)

```
            "adresa": ["Vukovarska", "111"],
            "email": "mmarulic@gmail.com"
    }
}
```

Primjer 2: XML

```
<osoba>
  <ime>Marko</ime>
  <prezime>Marulić</prezime>
  <OIB>55554444222</OIB>
  <rođendan>
          <dan>23</dan>
          <mjesec>04</mjesec>
          <godina>2180</godina>
  </rođendan>
  <mjesto>Split</mjesto>
  <adresa>
          <ulica>Vukovarska</ulica>
          <broj>111</broj>
  </adresa>
  <email>mmarulic@gmail.com</email>
</osoba>
```

5.3 Prednosti aktualnih formata za razmjenu

Danas su XML i JSON najčešće korišteni formati za razmjenu podataka na webu. CSV se katkad koristi za direktno preuzimanje strogo tabličnih podataka, čak i kada to nije najprimjereniji oblik (na primjer u podacima preuzetim s WoS i Scopus citatnih indeksa). Analizirajući JSON i XML neizbježno je uočiti prednosti aktualnih formata za razmjenu, obradu i zapis podataka. Za početak, oboje imaju logično definiranu sintaksu, tj. strukturu, čineći ih izrazito čitljivima računalima i ljudima. Suvremeni sustavi poput baza podataka i *online* aplikacija te programski jezici stoga znaju direktno raditi s podacima zapisanim u ovim formatima.

Oba formata nemaju unaprijed propisanu semantiku, već su otvoreni za slobodno imenovanje oznaka ili ključeva. XML ima jasno definirane standarde za propis semantike, dok je JSON potpuno otvoren i to ostavlja na brigu individualnim sustavima Time se ubrzava i olakšava proces zapisivanja podataka u formatu koji izaberemo te zbog toga oznake/ključevi budu prilagođeni prirodnom jeziku. U mnogim slučajevima nije potrebno imati znanje o tome koja identifikacijska oznaka označava što, već je vrlo jednostavno prepoznavanje podataka budući da se nazivi polja uglavnom ne kodiraju. U kompleksnijim slučajevima, naravno, još uvijek je potrebno pogledati specifikaciju kako bi se utvrdio jasan opseg polja (poput „naslov“ i „autor“). I JSON i XML su nastali s ciljem lakšeg prijenosa podataka i omogućavanje interoperabilnosti sustava. Nepropisana semantika formata omogućava i zapisivanje podataka koji su zapisani prema unaprijed definiranim standardima, pošto je potrebno samo da se nazivi

polja/oznaka/ključeva dogovore kako bi sistem koji prima/šalje podatke bio upoznat s istom. Drugim riječima, ovim formatima je jasno moguće zapisati, na primjer, i podatke prema MARC standardu, ali više ne u MARC formatu.

I JSON i XML (kao i HTML) su običan tekst koji se koristi kako bi server i preglednik mogli komunicirati bez dodatnih aplikacija. Ovakva metoda osigurava da će se podaci nesmetano razmijeniti bez opasnosti da format nije kompatibilan sa sustavima za prijenos podataka ili pak platformama koje šalje i primaju. Nekompatibilnost sustava bi moglo ugroziti strukturu podataka, ali i same podatke, pošto je moguće da sustav preinači podatke koje preuzima. Spomenuti formati ne ovise o specifičnoj vrsti softvera. Sve navedeno ih čini idealnim za razmjenu podataka.

Ne smije se zanemariti ni činjenica da programatska web sučelja koriste jedan od ova dva, a često se mogu koristiti i oba formata te su individualni korisnici, tvrtke i sustavi sve više navikli na podatke zapisane u takvim formatima. Korisnici vrlo jednostavno mogu obraditi podatke zapisane u XML-u ili u JSON-u u mnogim suvremenim tehnologijama. Različite baze podataka pružaju mogućnost preuzimanja baza podataka s čistim podacima koje onda korisnici mogu analizirati i izlučiti podatke koje im trebaju. S druge strane, mnoge baze podataka nude direktan zapis ovakvih podataka pa čak i relacijski sustavi (npr. PostgreSQL) čime zapravo postaju hibridni. Podatke možemo obrađivati bilo da su u fokusu pojedinačni podaci ili da ih promatramo kao skupinu. Kao praktičan primjer može poslužiti automatsko pobiranje podataka putem kakvog bibliografskog identifikatora (npr. ISBN i DOI) i zatim automatska izrada bibliografije ili održavanje vlastite zbirke tekstova na određenu temu. Kada su podaci dostupni u MARC formatu, ovaj problem iziskuje puno posla i programerskog znanja. Kada su podaci u kakvom jednostavnom JSON-u, ovaj problem je rješiv sa osnovnim programerskim znanjem odnosno na početničkoj razini.

Što se same računalne efikasnosti tiče, uspoređujući efektivnost JSON-a i XML-a u razmjenu podataka, ispostavilo se da podaci zapisani u JSON-u prosječno nadmašuje brzinu prijenosa podataka zapisanih u XML-u. JSON je jednostavan za analizu računalima iako koristi korisnikov procesor više no XML. Glavna prednost ovdje je da ne opterećuje poslužitelj, već klijenta. Memoriju koriste oba formata podjednako. Bitno je istaknuti da je vrijeme prijenosa podataka u JSON-u i XML-u podjednako bez obzira na količinu podataka.[80]

[80] Usp. Nurseitov, Nurzhan. Paulson, Michael. Reynolds, Randall. Izurieta, Clemente. Comparison of JSON and XML Data Interchange Formats: A Case Study. // Caine 9, (2009.). Str. 4. URL: https://www.cs.montana.edu/izurieta/pubs/caine2009.pdf (2018-05-26)

Prednosti ova dva formata ne odnose se samo na prijenos podataka, već i na arhiviranje. Pošto su oba obični tekst arhiviranje podataka u tim formatima je prigodno pošto ne ovise ni o jednoj posebnoj aplikaciji. Osim toga, samo razumijevanje podataka je moguće bez obzira na vrijeme pošto su nazivi oznaka zapisane prirodnim jezikom, uglavnom vrlo jednostavne. Postoje neslužbene preporuke da je XML bolji za arhiviranje od JSON-a. Preporuke su proizašle zbog činjenice da je XML prikladniji za zapisivanje podataka o dokumentima pošto omogućuje zapisivanje metapodataka i u obliku atributa, korištenje prostora za imenovanje te ima jasno definirane standarde za validaciju dokumenata i transformacije. Uz to, XML se, za razliku od JSON, može i koristiti za označavanje tekstova pisanih prirodnim jezicima što ga čini posebno korisnim u području računalne obrade prirodnog jezika.

5.4 Model podataka

XML i JSON propisuju strukturu i sintaksu**,** ne semantiku**.** Ukoliko želimo konzistentno zapisivati i razmjenjivati podatke za neku specifičnu domenu (npr. bibliografski podaci) potreban nam je standard**, odnosno shema,** koji propisuje postojanje, vrstu i opseg „polja“. Ovakav standard često nazivamo modelom podataka i upravo će se taj izraz koristiti u ovom radu.

Drugim riječima, model podataka je prvi korak koji se odlučuje prilikom osmišljavanja baza usmjerenih na objekte. Modelom se određuju odnosi među entitetima i atributima te određuje imena elemenata, tj. kako ćemo neki pojam označavati i značenje tih entiteta. Modeli se temelje na semantički dio. U knjižničarskoj zajednici postoje nekolicina modela, modeli za zapisivanje metapodataka, kao MARC XML, MODS, Dublin Core, su nasuprot tradicionalnih (MARC) primjetno fleksibilnije. Značajno je kod knjižničarskih modela da propisuju osim sintakse i detaljno semantiku i zahtjeva se od svih novo razvijenih modela da slijede tu praksu. Propisani modeli imaju ili linearni popis elemenata ili im je struktura hijerarhijska.[81]

Kao temelje moderniziranih modela ističu se: „modularnost (mogućnost kombiniranja elemenata iz različitih modela, vokabulara tako da su interoperabilni sintatički i semantički); proširivost (moguće je temeljne elemente ponavljati ili dodavati nove kako bi se prilagodilo lokalnim potrebama); preciznost (osigurava detaljno specificiranje značenja elemenata, klasifikaciju standardnih elemenata i definira raspon vrijednosti za određeni element); višejezičnost (omogućava modele podataka i zapisa na različite jezike i s prikladnim skupine

[81] Usp. Xie, I., Matiusak, K. Discover digital libraries: Theory and practice. Elsevier, 2016. Str. 133.

znakova).“[82] Modeli nastali za zapisivanje metapodataka su u većini slučajeva nastali sa svrhom opisivanja digitalnih objekata. Rijetko, skoro nikada se ne koriste za tradicionalnu građu iako su nastale na temeljima tradicionalnih formata.

Tijekom implementiranja novih modela je potrebno osmisliti detaljnu semantiku i sintaksu modela. Treba se postaviti pitanje da li je moguće zapisivati podatke koji dolaze iz takvog okruženja u formate koji skoro pa i nemaju značajno opisana pravila. Postoje primjeri, ali i dalje se zadržala temeljna struktura knjižničarskog formata. Činjenica da su se MARC pravila mogla samo prebaciti u XML ukazuje na mogućnosti koje nam otvaraju novi formati.

Formati kao JSON i XML time što nemaju propisanu semantiku omogućavaju implementiranje iste. U tim slučajevima se jednostavno standardizira model, ali je potrebno iskoristiti prednosti koje nam osiguravaju ovi formati. Modele koje možemo iskoristiti mogu biti ili usmjereni na univerzalne mogućnosti (RDF) ili usmjereni i specijalizirani na određenu domenu (BibJSON). Primjer uspješnog implementiranja semantike da se omogući umrežavanje podataka koristeći RDF-ovu metodu tripleta ukazuje na mogućnosti implementiranja i dogovaranja i drugih modela za zapisivanje podataka.

5.5 Umreženi podaci u JSON-u i XML-u

Usprkos činjenici da se nalazimo u dobu kada je pristup svim vrstama informacija izrazito pojednostavljen, još uvijek se nije iskoristio cijeli potencijal weba. Umreženost podataka je izrazito bitna u razvoju globalne mreže znanja. Već postoje različite metode kojima se osigurava jednostavna razmjena i komunikacija među sustavima, ali teško je postignuti visoku umreženost podataka. Ovo je posebno problem kada se uzme u obzir da računala obrađuju, analiziraju i pretražuju veliki broj informacija bez ljudskog uplitanja, ali za umrežavanje podataka i dalje su potrebni ljudi. JSON i XML kao formati su oba jednostavna za zapisivanje podataka, čitljivi i računalima i ljudima, a zapisi su jednostavno razmjenjivi jer su obični tekst.

Da bi se omogućilo umrežavanje podataka u oba formata postoje standardizirana proširenja spomenutih formata. Za JSON je to JSON-LD, a za XML je XML *Linking Language* (XLink). Oba standarda omogućuju stvaranje i opisivanje veza između izvora. Osim veza između izvora, XMLink omogućava i pripisivanje metapodataka linkova i izraženih veza koje se nalaze na odvojenoj lokaciji od umreženih izvora. Oznake za poveznice trebaju biti prepoznate pomoću XMLink aplikacija kako bi s njima moglo upravljati, a koristi se metodom

[82] Xie, I., Matiusak, K. Discover digital libraries: Theory and practice. Elsevier, 2016. Str. 133.

zapisivanja oznake u prostoru za naziv.[83] Drugim riječima, ime oznake za poveznicu može biti svaki put drugačije, nasuprot HTML-a gdje poveznice označava isključivo oznaka <a>.

Omogućava dvije vrste poveznica, proširene i jednostavne. Proširene omogućavaju sve funkcije XMLink-a, ali kao rezultat struktura im može biti kompleksna, pošto može uključivati elemente za vanjske i lokalne izvore, elemente za specificiranje izvora čitljivih ljudima i računalima, elemente za prelazak iz resursa u resurs. Jednostavni linkovi nude manje funkcionalnosti od proširenih pa nemaju posebno kompleksnu unutarnju strukturu, njihova svrha je da budu brze veze za ekvivalentnu proširenu vezu. Omogućuju jednostavne sintakse za izlazne veze samo za dva izvora. Jednostavne veze su podskup proširenih veza i sintaktički su različite.[84]

Nasuprot XMLink formata, JSON-LD je sintaksa za implementiranje mogućnosti Umreženih podataka u JSON. Njime je omogućeno da se JSON interpretira kao Umreženi podaci s minimalnim promjenama. Može se koristiti kao RDF, ali je dizajniran sa svrhom unapređivanja JSON-a tako da nije prijeko potrebno predznanje o RDF-u.[85]

Uz JSON elemente uveden je mehanizam za univerzalnu identifikaciju JSON objekata pomoću Internacionalnih identifikatora izvora (IRI, eng. *Internationalized Resource Identifier*) koji su ažurirani IRI-i, način razdvajanja ključeva (odnosno „polja“)podijeljenih između različitih JSON dokumenata tako da ih se mapira pomoću konteksta u IRI-e. Uveo se mehanizam u kojem se vrijednosti JSON objekta iz jednog izvora mogu referirati na JSON objekte u drugom izvoru na webu, omogućeno je bilježenje niza znakova sa svojim jezikom, način referiranja vrsta podataka kao datum i vrijeme i objekt za izradu jednog ili više RDF grafa u jednom dokumentu.[86]

Drugim riječima, JSON-LD omogućava različitim aplikacijama da komuniciraju skraćenicama, a da zbog razumijevanja konteksta podataka ne gube preciznost. S JSON-LD je moguće izraziti semantičko označavanje podataka i u drugim formatima za Umrežene podatke, zbog toga ga se i može koristiti za zapis podataka prema RDF modelu.

[83]Usp. XML Linking Language (XLink) Version 1.1. // W3C Recommendation. 2010. URL: https://www.w3.org/TR/xlink/#concepts (2018-05-24)
[84] Usp. Isto.
[85] Usp. JSON-LD 1.0. // W3C Recommendation. 2014. URL: https://www.w3.org/TR/json-ld/#introduction (2018-05-24)
[86] Usp. Isto.

6 Razmjena bibliografskih podataka

Ideja o razmjeni bibliografskih podataka nije nova, ali usprkos tome što su u počecima razvoja tehnologija knjižnice bile ukorak i slijedile napredak, u trenutku kada je web postajao globalna mreža podataka formati bibliografskih podataka su ostali u svojim početnim oblicima i nepovezani.

Premda je cilj svih knjižničnih formata i standarda razmjena i interoperabilnost, sve to prestaje u trenutku kada bi se trebala omogućiti ista ta interoperabilnost i razmjena sa sustavima i korisnicima u suvremenom umreženom računalnom okruženju. Sučelja na webu svoje podatke pružaju u JSON ili XML formatu, a najčešće u oba. Formati za razmjenu podataka na webu su jednostavni za zapisivanje te je omogućena lakša razmjena i interoperabilnost podataka na webu. Ne samo da su JSON i XML mogući za čitanje i pisanje ljudima, već su jednostavni za čitanje i računalima. Za razliku od MARC formata, na primjer, gotov svaki suvremeni programski jezik ima ugrađene mogućnosti čitanja ovih formata. To međutim ne rješava problem semantike dokumenta. MARC, kao što je opisano kasnije, nije problem direktno zapisati u XML. Ipak, osobe i računalo ako nemaju u dodatnog znanja o tome što koje polje u MARC formatu označava teško mogu razumjeti zapis. Naravno, osigurano je da se podaci prikažu u katalogu u obliku tradicionalnih kataložnih zapisa, napisanih prema ISBD standardu i svim njegovim interpunkcijama, ali ako su nam potrebni čisti podaci (na primjer za potrebe direktnog citiranja u znanstvenoj literaturi, prikupljanje te literature ili bibliometrijsko istraživanje literature), skoro su nečitljivi.

Dodatan problem su sustavi kakvi se koriste u praksi. U slučaju online baza podataka vrlo često postoji mogućnost preuzimanja većih količina bibliografskih zapisa istovremeno i automatski, dok u katalozima te opcije u načelu nema. Da bi se moglo preuzeti veći broj bibliografski zapisa iz kataloga, trebamo ući u svaki zapis zasebno i ručno kopirati podatke. Bibliografski zapisi preuzeti na taj način su tekstualni, softveri koje korisnici možda koriste za izgradnju svojih zbirki, kao Zotero, Mendeley, nemaju opciju za obradu takvih podataka. Ne samo da softveri ne mogu obraditi podatke, već se u podacima kada ih ručno preuzimamo obuhvate i nepotrebni podatke kao nazivi ISBD polja i interpunkcije.

Mogućnost direktne razmjene podataka s korisnicima je onemogućena jer se tradicionalno nije razvijala. Korisnici moraju postaviti upit kako bi im knjižničari napravili popis literature u tekstualnom obliku koji nisu u mogućnosti obraditi računalno na strukturiranoj razini. Naravno, postoje razlike u korisnicima i nemaju svi korisnici istu potrebu

za obradu i analizu podataka. Postoji i razlika s kojim ciljem korisnici koriste katalog, da li zbog istraživanja ili kako bi doznali za literaturu s kojom bi mogli ispuniti svoje slobodno vrijeme. Postoji i razlika u vlastitim zbirkama koje korisnici izgrađuju gdje u nekim slučajevima korisnicima dovoljan popis naslova i autora, a u drugima imaju ozbiljnu potrebu za upravljanjem većim skupom metapodataka i to posebno kada prate tematska područja koja ih zanimaju. Knjižnice bi bez obzira, kao centri informacija, trebali omogućiti da se njihovi podaci mogu preuzeti u aktualnim formatima za razmjenu. Ovo bi omogućilo i bogatije sustave u smislu API komunikacije, mogućnosti preporuke i istraživanja korisnika i slično.

Podaci okupljeni katalogizacijom su bogati i provjereni, osiguravanjem da su podaci dostupni u aktualnim formatima, ne da bi se samo omogućila lakša razmjena, već bi i pristup tim podacima bio jednostavniji. Time što se pruže podaci u drugim formatima ne znači nužno i gubitak strukture te ako već postoji strah od pružanja podataka javno može se naučiti taktike baza podataka u pružanju bibliografskih podataka. Iako se preferira zadržavanje MARC formata za zapis podataka, postoje određeni pokušaji da je MARC zapisa omogućen i u novim aktualnim formatima. Previše je do sada uloženo u infrastrukturu da bi se MARC jednostavno napustio, ali pri tome ne treba zaboraviti već spomenute probleme u paradigmi. Jedina su, stoga, dugoročna rješenja modernizirati MARC format ili ga zamijeniti nečim novim što će prenijeti već bogato znanje uloženo u MARC, ali i omogućiti suvremene koncepte nastale razvojem podataka u umreženom svijetu.

6.1 Ograničenja u razmjeni i interoperabilnosti bibliografskih podataka

Razmjena u knjižničarskim ustanovama i sustavima koji se temelje na ISBD standardu je uspješna jer ovako stvoreni zapisi su interoperabilni, ali interoperabilnost sa sustavima izvan knjižničarskih krugova je neuspješna i često nepostojana. Problem je što to uključuje korisnike koji sve češće žele preuzimati strukturirani zapis, a ne onaj već formiran u tekst prikladan ljudskom oku. U slučajevima kada se podaci žele ili moraju dijeliti, teži se zadržavanju knjižničarske forme zapisivanja podataka, koji su često nerazumljivi korisnicima, a ponekad i neupotrebljivi izvan knjižničarske zajednice. Srž problema leži u tome što je ISBD dizajniran za isključivo ljudsku, a ne računalnu iskoristivost. Taj problem je posebno došao do izražaja kada se ISBD trebao reorganizirati u skladu s pravilima Semantičkog weba i FRBR modela. Također, znanstvena zajednica (koja ima čestu potrebu razmjenjivati bibliografske podatke) razvija vlastite jednostavne formate za razmjenu koje knjižnice još uvijek rijetko pružaju svojim korisnicima (npr. format BibTeX) i koji su ostavština softvera za upravljanje referencama,

BibTeX je, na primjer, kao softver dio LaTeX sustava izrade dokumenata koji po prirodi podrazumijeva baratanje većim brojem referenci. S vremenom, BibTeX kao format koji je taj softver koristio postaje *de facto* standard za razmjenu bibliografskih metapodataka u znanstvenoj zajednici.

Način pristupa knjižničarske zajednice novim alatima i pomagalima se može vidjeli kroz sve etape novih tehnologija. Iako su se prednosti za računalno čitljive kataloge vrlo brzo prepoznate, način na koji su se bibliografski podaci prebacili je taj da su se zadržali elementi, interpunkcije i izgled kataložnog listića i samo se prešlo iz analognog u digitalni format bez promjene paradigme. Posebno je vidljiv otpor za promjenu u nedavnim pokušajima tzv. FRBR-iziranja kataloga koji uvodi relacije u ovaj model čime knjižničarski metapodaci postaju adekvatniji za pohranu u relacijski model.

U smislu podataka strukturiranih za računalnu iskoristivost, zanimljivo je napomenuti da je MARC nastao prije relacijskih baza podataka, odnosno ideje koja je obilježila shvaćanje računalnih podataka. Pojava FRBR-a je upravo reakcija knjižnične zajednice na relacijsko shvaćanje podataka. U međuvremenu se u svijetu baza podataka sve više spominje objektni pristup, takozvani NoSQL pristup koji je najčešće pohrana JSON podataka te mrežne baze podataka koje se temelje na teoriji grafova čime zapravo spajaju objektni i relacijski pristup.

Pokušaji da se bibliografski zapis prilagodi suvremenim trendovima zapisivanja podataka naišli su na veliki otpor, najviše stoga jer se smatra da se prebacivanjem gubi struktura i redoslijed koju nam je ISBD pružio.[87] „Pod nedostatkom strukture se smatra odsutnost hijerarhijskog prikaza predmeta i gubitak konteksta djela pošto svako novo izdanje, prijevod i preradba djela dobije svoj zasebni zapis. Uvođenjem FRBR-a moglo bi se vratiti dio strukture i pružiti grafičku reprezentaciju konteksta za svaki novi zapis. Ali postavlja se pitanje da li je to uopće nešto što korisnika interesira."[88] Osim toga, identifikacija redoslijedom je nešto što se u novim sustavima s razlogom napušta jer redoslijed može biti stvoren po potrebi. Također nepovezanost zapisa o istom naslovu je problem sustava, a ne problem tih zapisa.

[87] Elaine Svenonius, The Intellectual Foundation of Information Organization. Cambridge, MA: MIT Press, 200. Str. 62. Citirano prema: Bianchini, Carlo; Guerrini, Mauro. From Bibliographic Models to Cataloging Rules: Remarks on FRBR, ICP, ISBD, and RDA and the Relationships Between Them. // Cataloging & Classification Quarterly 47, 2 (2009.). Str. 108. URL: http://eprints.rclis.org/16303/1/CCQ_Bianchini-Guerrini_From_907343194-1.pdf (2018-05-03)

[88] Le Boeuf, Patrick. FRBR: Hype or cure-all? Introduction. // Cataloging & Classification Quarterly 39, 3-4 (2005.). Str. 4.

Istaknuto je da se knjižničarski standardi neće napuštati dok se novi, kao što je FRBR, ne testiraju do kraja. Smatra se ISBD posebno važnim zadržati pošto je jedan od najraširenijih međunarodno prihvaćenih standarda. Iako se ISBD prilagođavao s dolaskom nove građe, ali postavlja se pitanje da li ISBD zadovoljava potrebe današnjih publikacija i posebno iskoristivosti za elektroničke publikacije i organizaciju građe u digitalnoj sredini. Manje je slučajeva u kojima nam ISBD može pomoći u prenošenju bibliografskih zapisa u elektroničke formate, kada se podaci sve više proizvode u elektroničkom formatu.[89] Upravo je zapravo obratno istina. ISBD bi se trebao moći proizvesti na zahtjev, a ne biti glavna paradigma modela za pohranu podataka.

Postoji napredak u interoperabilnosti bibliografski zapisa, ali ako se gleda sam ispis podataka korisnicima, struktura je ista kao i kod kataložnih listića. Ako se korisnici i odluče koristiti čiste MARC podatke koji nisu pročišćeni za prikaz prvi problem moraju utvrditi koju MARC verziju knjižnica koristi i da li postoje kakve lokalne osebujnosti u zapisima. Također, MARC se koristi samo u knjižnicama i za njega gotovo da ne postoje korisnicima dostupna rješenja za računalno upravljanje ovim zapisima što poražava svrhu izvan posve profesionalnog korištenja.

Navedeno postaje sve problematičnije jer digitalna sredina stavlja korisnika u poziciju organizatora vlastite zbirke pa su korisnici često u poziciji da zbog osebujnosti zapisa bibliografske informacije kreću dohvaćati iz online usluga koje nisu uopće vezane za knjižnične ustanove. Već se kod MARC-ova može primijetiti način na koji se bibliografskih podaci i struktura kojom su zapisani prebacuje iz formata u format. Izvori odakle se dobiju bibliografski podaci možda jesu drugačiji, ali sam izgled podataka i način na koji su zapisani (interpunkcije i redoslijed elemenata) ostaje isti.

Tennant i Andresen ističu kao ključne nedostatke MARC formata nefleksibilnost i nisku granulaciju. Ističe se kako je MARC dogovoren prema kartičnom katalogu što može stvarati probleme pri opisivanju digitalnih, vizualnih i multimedijskih izvora te da promjene koje se mogu dodati na lokalnoj razini zbog nefleksibilnosti ometaju razmjenu između lokalnih i skupnih kataloga. Za granulaciju ističu također da je problem što je MARC dogovoren prema

[89] Usp. Bianchini, Carlo; Guerrini, Mauro. From Bibliographic Models to Cataloging Rules: Remarks on FRBR, ICP, ISBD, and RDA and the Relationships Between Them. // Cataloging & Classification Quarterly 47, 2 (2009.). URL: http://eprints.rclis.org/16303/1/CCQ_Bianchini-Guerrini_From_907343194-1.pdf (2018-05-03)

kartičnom katalogu gdje su u centru dokumenti te je skoro nemoguće kvalitetno izraziti hijerarhijske odnose među podacima.[90]

Kao što u samom priručniku piše "osnovni cilj UNIMARC-a jest promicanje međunarodne razmjene bibliografskih podataka u strojno čitljivom obliku između nacionalnih bibliografskih službi." Iako je cilj promicanje razmjene bibliografskih podataka, ta razmjena se temelji isključivo na knjižnice i njihove potrebe, zanemarujući ostale blokove mogućih korisnika bibliografskih podataka. Osim toga, i MARC21 i UNIMARC se promoviraju kao formati za međunarodnu razmjenu, ali niti jedna MARC verzija nije dominantna, već svaku MARC verziju koriste određeni broj država.[91]

Također, različiti integrirani knjižnični sustavi mogu unijeti svoje osebujnosti u inače isti standard. Kao što je već prije navedeno, MARC je osmišljen kako bi se olakšala, ubrzala i financijski olakšala razmjena bibliografskih podataka među knjižnicama. No međutim, događa se da knjižnice koriste različite softvere koji im olakšavaju zapisivanje u MARC formatu, a različiti softveri ne samo da često koriste drugačiji MARC format već često imaju i vlastite osebujnosti koje unose u MARC zapise kojima upravljaju. Samo se u Hrvatskoj koristi najmanje 5 različitih, službenih softverskih rješenja za knjižnično upravljanje MARC zapisima.

Pored razlika koje je unio sam softver, često se događa da knjižnice same iskoriste određena neiskorištena polja u nestandardne svrhe što u slučaju pokušaja komunikacije s drugim knjižnicama može dovesti do problema. Ako sama struka ne može riješiti interoperabilnost zapisa za vlastite potrebe postavlja se pitanje koliko je ta struka u poziciji da se nastavi nazivati "informacijskim stručnjacima" i da razvije kvalitetna rješenja za potrebe korisnika. U 21. stoljeću postaje razvidno da se korisnici za potrebe pobiranja bibliografskih podataka okreću uslugama poput Google Books/Scholar, Amazon *online* trgovine ili pak bazama znanstvenih tekstova\bibliografskih zapisa koje uglavnom nemaju direktne veze s knjižničarskom djelatnošću.

Nedostatak je i također u tome što korisnici teže mogu okupiti podatke koji su zapisani u MARC formatu, nego što su u nekom komercijalnim bazi podataka. Gotova korisnička

90 Usp. Gonzales, Brighid M. Linking libraries to the web: linked data and the future of the bibliographic record. // Information Technology and Libraries (Online) 33, 4 (2014). Str: 11. URL: https://ejournals.bc.edu/ojs/index.php/ital/article/download/5631/pdf (2018-10-07)

91 Leif Andresen. After MARC—What Then? // Library Hi Tech 22, 1 (2004). Str. 41. Citirano prema: Gonzales, Brighid M. Linking libraries to the web: linked data and the future of the bibliographic record. // Information Technology and Libraries (Online) 33, 4 (2014). Str: 11. URL: https://ejournals.bc.edu/ojs/index.php/ital/article/download/5631/pdf (2018-10-07)

softverska rješenja čak i ako dopuštaju uvoz MARC-a, rade samo s određenim formatima MARC-a i to iz određenih knjižničnih softvera tako da su u praksi ti zapisi često neupotrebljivi za razmjenu. Programski jezici ne mogu čitati MARC format bez posebnih proširenja koja su rijetka i gotovo nikad ne pokrivaju cijelu porodicu MARC formata. Ta proširenja, ako i postoje, imaju isti problem kao i u slučaju gotovih korisničkih softverskih rješenja. Drugim riječima, s MARC formatom je danas nepotrebno otežano raditi programatski, iako mu je to primarna namjena. Čak i ako korisnik uspije pristupiti MARC podacima na sintaktičkoj razini, ne može puno s njima bez bogatog poznavanja standarda i knjižnične prakse. Navedeno onemogućuje korištenje MARC-a izvan stručne knjižničarske sredine, a knjižničarska struka ne razvija vlastite računalne formate za komunikaciju s korisnicima. MARC format bi se mogao koristiti kao glavni format iz kojeg se proizvode jednostavniji formati poput npr. BibTeXa, ali sama priroda MARC formata (granularnost i orijentiranost na redoslijed i interpunkciju s kataložnih listića) ovo znatno otežava. Čest je slučaj, na primjer, da ISBD interpunkcija postane dio naslova te da se kojekakvi podaci o odgovornosti koji nisu autorstvo pretvore u autorstvo.

Također kao i MARC 21, UNIMARC je podržan standardom ISO 2709 za strukturu i sustav označitelja sadržava strojno čitljivi zapisa. Dapače, navedeno je i najveći problem MARC formata za suvremeno doba: on služi elektroničkoj pohrani informacija za izradu kataložnog listića, a ne kao značajan prijelaz u paradigmu upravljanja (bibliografskim) podacima računalima.

Danas postoje različite metapodatkovne sheme koje su odgovor na nove podatke koji se smatraju nužnim za opis, a klasični knjižničarski standardi ih ne mogu ukomponirat u svoje strukturirane zapise. S novom građom došle su i nove uloge, a knjižničarski standardi ne prate promjene. Ne samo da su nastali s novom građom, već su i baze podataka standardizirale neke informacije koje su korisnicima izrazito korisne, a u knjižničarskim katalozima ih ne pronalazimo.

Jedan od primjera podataka je sažetak (*abstract*), u bazama podataka, ali i u novim metapodatkovne shemama. Sažetak pruža kratke informacije o čemu je primjerak. U knjižničarskom svijetu se te informacije pružaju oznakama neke od knjižničarskih klasifikacijskih sustava ili predmetnicama. Nijedan od ova dva načina označavanja sadržaja nije toliko detaljan kao što je sažetak. Osim sažetka postoji još podataka koje su prepoznate kao danas izrazito korisne, a da ih knjižničarski standardi ne osiguravaju u svojim zapisima. Čest slučaj su novi identifikatori poput DOI-a. To je samo primjer kako trenutno standardi u nekim slučajevima ne slijede potrebe svojih korisnika.

6.2 Bibliografski podaci i suvremeni formati za razmjenu

Sve većim rastom digitalnih izvora i digitalnih zbirki, baštinske ustanove (arhivi, knjižnice i muzeji) su uočili potrebu za fleksibilnijim, interoperabilnijim i detaljnijim opisom, ali i upravljanjem te razmjenom ovako nastalih podataka. Za svrhu opisivanja digitalnih objekata osmišljene su različite metapodatkovne sheme, kao što je Dublin Core, *Visual Resources association core* (VRA), *Encoded archival description* (EAD), METS, MODS i mnogi drugi. Neke od ovih su nastali kao reducirani bibliografski opisi za potrebe weba, a druge kako bi podruprijele zapisivanje MARC podataka u nove formate.

Iako su osmišljene na uzor tradicionalnih formata, stručnjaci su uvidjeli određene prednosti univerzalnih formata nastalih izvan knjižničarske zajednice kao XML. Naravno, i kod metapodatkovnih shema postoje problemi. Struktura je i dalje ili linearna ili hijerarhijska, ali su osmišljene s ciljem interoperabilnosti i suradnje s vanjskim suradnicima. Suradnja se odvija i s korisnicima koji u nekim slučajevima mogu dodavati svoje oznake/predmetnice/tagove ili s autorima/izdavačima koji bi pružili metapodatke uz svoje djelo. Interoperabilnost zapisa ovih shema se odvija na sintatičkoj razini, rijetko na semantičkoj. Nažalost, metapodatkovne sheme su osmišljene s ciljem pružanja opisa za digitalne objekte u digitalnim knjižnicama te se primjena ovih shema ne koristi i za izvore opisane u MARC formatu, usprkos tome što neke sheme nastale upravo prema MARC formatu (MODS i METS). Iako su ove sheme interoperabilnije i prilagođenije web prostoru, u većini slučajeva se pruži mogućnost preuzimanja MARC zapisa u MODS ili METS u kojima se MARC zapis mapira bez ikakvih gubitaka. Prilikom adaptacija MARC zapisa u druge formate se teži da se zadrži MARC struktura te se samo prebaci sintaksa i podaci u drugi format bez preinaka.[92]

Drugim riječima, iako je MARC podatke u suvremenim formatima za razmjenu lakše čitati i mijenjati suvremenim tehnologijama, podložna paradigma je još uvijek kataložni listić što bi trebalo nanovo osmisliti za umreženo digitalno doba. Oznake se nazivaju prema nazivima polja i potpolja, ugniježđeni podaci su neprikladno upisani i nastavljaju se koristiti ISBD interpunkcijski znakovi kao dio vrijednosti polja radije no da se automatski dodaju posebnim pravilima. Tijekom prebacivanja se ne iskoristi potpuni potencijal suvremenih formata. Na primjer polje autor, tj. polje za podatak o odgovornosti, umjesto da se autorovo ima odvoji na elemente ime i prezime kako bi se još detaljnije precizirali podaci zapiše se kao kako bi u MARC zapisali.

[92] Usp. Xie, I., Matiusak, K. Discover digital libraries: Theory and practice. Elsevier, 2016. Str. 129-143.

Korištenje aktualnih formata za zapisivanje podataka ne znači nužno i potpuno odbacivanje knjižničarskih standarda i formata, ISBD i MARC, već da se omogući pristup bogatim informacijama s ciljem globalnog umrežavanja podataka. Ne samo da bi se web obogatio bibliografskim podacima i omogućilo lakše korištenje podataka korisnicima, već bi se mogli i bibliografski zapisi obogatiti i proširiti. Važno je napomenuti kako neke usluge već imaju dobro riješeno preuzimanje bibliografskih metapodataka u suvremenim formatima poput Crossrefa i raznih baza podataka. Ipak, ovi primjeri dolaze izvan knjižnične zajednice. Podaci o časopisima pogotovo bi imali beneficije dodavanjem Digitalnih identifikatora objekta (DOI) za svaki zasebni članak, pošto se u katalogu časopisi opisuju kao jedinka i upravo je dodavanje novih identifikatora i veza jedan od primjera gdje ugrađena proširivost formata poput XML-a i JSON-a pokazuje svoje snage.

Modeli i sheme koji su nastali prema MARC formatu, s ciljem da se omogući u aktualnim formatima su: MARCXML, MODS i BibJSON.

6.2.1 MARC XML

„MARC XML je okvir za rad s MARC podacima u XML okruženu. Okvir je osmišljen tako da bude fleksibilan i proširiv kako bi korisnici mogli prilagoditi rad s MARC podacima svojim potrebama. MARC XML sadrži sheme, stilove i softverske alate koje je razvila i održava Kongresna knjižnica.“[93]

MARC XML je XML shema koja se može koristiti za potpunu reprezentaciju MARC zapisa u XML-u, kao nastavak sheme za Standard za kodiranje i prijenos metapodataka (*Metadata Encoding & Transmission Standard* (METS)). METS je standard koji se koristi za kodiranje opisnih, administrativnih i strukturalnih metapodataka koji se odnose na objekte unutar digitalnih knjižnica, izraženi XML shemom. Dalje, MARC XML omogućava reprezentaciju metapodataka za Inicijativu otvorenih arhiva (OAI), za opisa zapisa u XML sintaksi i za metapodatke u XML koji dolaze s elektroničkim izvorom. Jedna od prednost MARC XML-a je što podržava sve podatke zapisane u MARC-u bez obzira na format i to što je komponentno orijentirana. Drugim riječima, proširiva arhitektura omogućava korisnicima da dodaju različite softvere kako bi izgradili prilagođenu strukturu.[94]

[93] MARC XML: Uses and Features. // The Library of Congess. URL: https://www.loc.gov/standards/marcxml/marcxml-overview.html (2018-05-27)
[94] Usp. MARC XML Design Considerations. // The Library of Congess. URL: https://www.loc.gov/standards/marcxml/marcxml-overview.html (2018-05-27)

Ističe se kako je prijenos podataka bez gubitaka, što ukazuje na činjenicu da se MARC zapis samo prebacio u XML bez ikakvih preinaka i prilagodbe, samo prebacivanje s dodanim XML elementima.

Primjer 3: MARC XML

```
<datafield tag="260" ind1=" " ind2=" ">
  <subfield code="a">Santa Barbara, CA :</subfield>
  <subfield code="b">University of California, Santa Barbara, The Early Modern
                              Center,</subfield>
  <subfield code="c">2008.</subfield>
</datafield>
```

Kao što možemo vidjeti na primjeru svi elementi koje možemo pronaći u MARC zapisu se nalaze i u MARC XML zapisu. Ponovno korisnici koji dobiju ovakav zapis bez poznavanja MARC formata, koje polje označava koji podataka ili s posjedovanje alata koji bi konvertirao zapis, teško mogu išta napraviti s podacima. Osim toga, ne može se zanemariti i ISBD interpunkcija koje možemo pronaći u zapisu i neiskorištavanje mogućnosti ugnježđivanja podataka. Također bitno je napomenuti da je MARC XML nastao prema MARC 21, što znači da se okvir ne može koristiti i za podatke zapisane u UNIMARC formatu.

6.2.2 MODS

Shema za opis metapodatkovnih objekata (*Metadata Object Description Schema* (MODS)) je shema za bibliografske elemente, posjeduje slična svojstva kao i MARC XML s time da se ističe da je shema orjentiranija prema krajnjim korisnicima. Ograničenje je nemogućnost vraćanja zapisa u prvobitan oblik, tj. u MARC 21 format, jednom kada smo ih prebacili u MODS. Skupina elemenata je bogatija od Dublin Core skupine, elementi su kompatibilniji s knjižničnim podacima nego ONIX te je skupina elemenata jednostavnija od MARC formata. Postoji nacrt za elemente i atribute s ciljem lakšeg mapiranja MARC 21 zapisa u MODS.[95] Ponovno je format prilagođen samo za MARC 21 zapise.

6.2.3 BibJSON

BibJSON je prijedlog za metodu zapisivanja bibliografski podaci u JSON formatu. Iako samo prijedlog za sada je najbolje prilagođen primjer prebacivanja podataka iz MARC zapisa u neki drugi format. BibJSON ide korak dalje od MODS-a i MARC XML-a, ali i dalje postoje nedostaci i mogućnost dubljeg prikaza podataka od onoga što je predloženo.

[95] Usp. MODS: Uses and Features. // The Library of Congess. URL: https://www.loc.gov/standards/mods/mods-overview.html (2018-05-27)

Na primjer, ime autora nije odvojeno na elemente ime i prezime, već je listu vrijednosti (arrays). Osim ključa autor, listu vrijednosti (arrays) su još: urednik, licenca, identifikatori, linkovi, a časopisi su objekt. Prednost je utoliko što nema sve ISBD interpunkcijske znakove zapisane u podatke, već samo poneke kao zarez nakon autorova prezimena ili crtica između stranica. Specifična sintaksa još uvijek nije propisana tako da je moguće uvoditi promijene, a i postoje određeni prijedlozi kako raščlaniti BibJSON u neki drugi format poput CSV-a ili, specifičnije, RIS-a. Pogotovo je važno što se bibliografski podaci zapisani u BibJSON-u mogu lako pretvoriti u BibTeX zapise pošto su ključevi BibJSON-a temeljeni na BibTeX poljima. Omogućeno je i umrežavanje podataka korištenjem sintakse za zapis umreženih podataka u JSON-u, tj. JSON-LD.[96]

Primjer 4.: BibJSON

```
{
  "title": "Open Bibliography for Science, Technology and Medicine",
  "author":[
          {"name": "Richard Jones"},
          {"name": "Mark MacGillivray"},
          {"name": "Peter Murray-Rust"},
          {"name": "Jim Pitman"},
          {"name": "Peter Sefton"},
          {"name": "Ben O'Steen"},
          {"name": "William Waites"}
  ],
  "type": "article",
  "year": "2011",
  "journal": {"name": "Journal of Cheminformatics"},
  "link": [{"url":"http://www.jcheminf.com/content/3/1/47"}],
  "identifier": [{"type":"doi","id":"10.1186/1758-2946-3-47"}]
}
```

[96] Usp. How to do BibJSON. URL: http://okfnlabs.org/bibjson/#purpose (2018-05-30)

7 Zaključak: Dobra praksa u razmjeni bibliografskih metapodataka

Razmjena bibliografskih podataka među sustavima, uključujući knjižničarske i ostale, je moguća. Da bi se razmjena ostvarila potrebno je MARC zapis omogućiti i u drugim suvremenim formatima te razmisliti o modelu podataka koji se zapisuju. Postoje određeni problemi koje treba uzeti u obzir kada se razmatra prebacivanje zapisa. Kao što smo vidjeli kod MARCXML i MODS zapis se prebacuje bez ikakvih gubitaka, to znači da su elemenata u XML-u nazvani prema MARC 21 poljima i potpoljima. Treba se uzeti u obzir da je MARC format izrađen sukladno ISBD-u koji je izrazito strukturiran i standardiziran, ali ne s računalom u vidu. XML i JSON propisuju strukturu, ali nemaju sami po sebi propisane standarde za korištenje u specifičnim domenama, odnosno propise za podatke određene vrste (poput bibliografskih metapodataka).

Za uspješno zapisivanje bibliografskih podataka u aktualne standarde nije dovoljno samo konvertiranje iz jednog standarda u drugi, već je nužno razlučiti koji podaci su nužni da bi bili u zapisu i na kojoj razini. Samo određena skupina podataka je korisna korisnicima, a dodatne razine često korisne poput raščlanjivanja imena autora potencijalno do strukturne razine na kojoj se razlikuje ime od prezimena. To bi u većini slučajeva bili podaci koje bi dobio kada zatražio citiranu literaturu. Zbog toga je bitno da se podaci ne prebacuju iz jednog u drugi već promišljeno izdvoje i smisleno osmisli model podataka. Korisnika samo zasićujemo svim dodatnim podacima koji su potrebni za struku. Osim toga, potrebno je i imenovati elemente s prirodnim jezikom, a ne s oznakama polja. Kada bi ostavili nazive korisnik ponovno bez znanja o MARC-u ne bi mogao obraditi podatke.

Potrebno je osmisliti model kojim bi se propisala semantika i u nekim slučajevima sintaksa kako bi se osiguralo standardizirano zapisivanje i uspješno mapiranje elemenata u formatima koji su otvoreni za prilagođavanje. Prilikom pisanja modela treba se uzeti u obzir problemska polja i kako pravilno ugnježđivati podatke, a interpunkcijske znakove ne držati kao sadržaj polja već dodavati automatskim pravilnicima. Do sada prilikom prebacivanja podataka u nove modele nisu se uzimale u obzir mogućnosti koje ti modeli pružaju. Podaci bi se zapisali preciznije jer je vrijednosti moguće zapisivati na više razina granularnosti.

Na primjer entitet autora možemo odvojiti na atribute ime i prezime i ostale dodatne atribute sukladno s kulturom odakle potječe ime autora. Prednost odvajanja tog entiteta nije

samo za korisnike, već i za knjižničare. Kada je autorovo ime zapadnog podrijetla, u većini slučajeva nam je jednostavno napraviti distinkciju između imena i prezimena, ali različite kulture ima drugačiji način imenovanja. U slučajevima kada imamo imena iz kulture s drugačijim pravilima korištenje prirodnog jezika u imenovanju atributa jednostavnije nam je definiranje kako predstaviti ime u bilo kojoj situaciji. Navedeno je vrlo korisno u kasnijem automatskom baratanju tim podacima i to posebno prilikom automatske izrade bibliografija i citiranja literature. Treba napomenuti i da je dogovoreno da se u nekim slučajevima autor teksta ne piše u UNIMARC zapisu pod potpolje $f (Prvi podataka o odgovornosti), već se kao autora smatra ilustrator. Ta informacija je prikazana u prikazu pod Naslov. U ključ autor onda treba pisati ime, prezime, vrsta autora i moguće je dodati još dodatne elemente kao srednje ime koje može biti lista vrijednosti i tako dalje. Važno je također razlikovati autorstvo od ostalih podataka o odgovornosti što u knjižničarskim sustavima često navodi laike na krive zaključke.

Za neke atribute je moguće dogovoriti kodove, kao što već postoje kod za jezik, za zemlju i kod za vrstu autora, ali za neke je praktičnije da budu nepropisani kao kod za vrstu autora jer je danas teško pretpostaviti koja sve vrsta autorstva može biti, a već sada ako se treba koristit „više od jednog koda za jednog autora, pojedini se kodovi moraju unositi u zasebna potpolja $4.“[97]

Problem je osim toga i što se događaja neusklađivanje zapisa pa ponekad piše u potpolje da je autor ilustrator, crtač, strip. Nije samo problem kada je autor ilustrator, kada imamo autora da je poznat pod inicijalima kao što je autorica Harry Pottera ili autor Gospodara prstenova, tada se u potpolje piše ime: J. [John] R. [Ronald] R. [Reuel] Tolkin. Podaci su ovako i prikazani. Još jedan od problema je u polju 200 potpolje $g (Idući podaci o odgovornosti) koji je ponovljiv te se može dogoditi pojava da u jednom $g-u otvorimo uglatu zagradu ([ ), a zatvorimo je tek u drugom $g, kao u slučaju kada imamo dva prevoditelja. Kada bi takve podatke dobio softver podaci se ne bi prikazali kao zajednički, kao što je u slučaju kataloga, već bi dobili podatke s uglatim zagradama koje su višak. Kada bi prebacivali u suvremeni format to bi trebali napraviti bez uglatih zagrada, odnosno ne na razini teksa nego na razini strukturalnog razdvajanja informacija. Na primjer, u slučaju djela Gospodar prstenova prevoditelj je Zlatko Crnković, ali Neven Antičević preveo prepjev stihova te se oba zapisuju u polje o idućem podatku o odgovornosti.

[97] UNIMARC: bibliografski format / [autorizirani prijevod IFLA [i. e.] International Federation of Library Associatons and Institutions ; hrvatski prijevod formata iz 1999. osuvremenila Marijana Tomić; stručna redakcija prijevoda Mirna Willer]. Zagreb: Hrvatsko knjižničarsko društvo; Zadar: Sveučilište, 2009. Str. 137.

Pod element Autor u prikazu se nalazi ime i prezime autora što je proizašlo od polja 700 (Osobno ime - primarna odgovornost). Prevoditelj, osoba koja je napisala književni prijedlog ili koja je oslikala djelo se nalazi u prikazu pod Ostali autori, podaci zapisani pod polje 702 (Osobno ime - sekundarna odgovornost). Podaci u bloku 7 su detaljnije zapisani i precizirani i kada bi se podaci prebacivali u JSON format bilo bi prikladnije iz tih polja preuzimati podatke o autorima pošto su u polju 200 podaci popunjeni s dodatnim interpunkcijama i elementima koji bi samo nepotrebno zasićivali podatke. U slučaju da postoji više autora, kao kod znanstvenih radova ili zbornika, može se zanemariti entitet o ulozi, ali prednost je utoliko što se svi autori mogu zapisati pod jednim roditeljskim entitetom. Entitet autor dakle mora biti lista objekata. Također, ne smije se zaboraviti da se interpunkcije ne zapisuju kako bi bili podaci što „čišći".

Sljedeći problem je prikaz podataka u Naslov elementu. U prikazu zapisa, pod Naslov imamo podatke o naslovu, podatke o odgovornosti i različite interpunkcije. Dok je stručnjacima jasno zašto, laicima to ne odgovara konceptu „naslov". Nadalje, ovi podaci su opet strukturirani na razini običnog teksta za prikaz čovjeku radije no raščlanjeni na podatkovnoj razini. Svi podaci osim naslova su u ovom slučaju suvišni. Osim toga, sve interpunkcije i skraćenice (izd., ilustr.) se moraju upisivati ručno što dovodi do opasnosti da se dogode pogreške. Naslov treba razdvojiti tako da u naslovu samo piše naslov bez obzira o kojoj je vrsti riječ. Također ga možemo napraviti kao objekt u slučajevima kada imamo usporedni stvarni naslov ili originalni naslov, koji se nalaze u napomeni.

Podatak o izdanju se nalazi u polju 205 (Izdanje) te u slučajevima kada imamo samo podataka u $a (Podatak o izdanju) potrebno je napisati: 2 izd.. U JSON formatu se ne bi zapisivalo izd. već bi se samo naznačilo da je to entitet „edition". U slučaju da je u pitanju časopisi volumen i broj se zapisuju kao lista vrijednosti.

Primjer 5.: Prijedlog bibliografskog zapisa u JSON-u za potrebe korisnika.

```
{
	"source": {
		"type": "book",
		"set": "Harry Potter",
		"title": "Harry Potter i kamen mudraca",
		"original title": "Harry Potter and the philosopher's stone"
	},
	"responsibility": [
		{
			"type": "author",
			"given": "Joanne",
			"middle": "Kathleen",
			"family": "Rowling"
		},
```

```
                {
                        "type": "translator",
                        "given": "Zlatko",
                        "family": "Crnković"
                }
        ],
        "edition": 7,
        "imprint": {
                        "city": "Zagreb",
                        "publisher": "Algoritam",
                        "year": 2002
        },
        "language":"hr"
}
```

Ovo je samo primjer koji bi se elementi mogli nalaziti u jednom prikazu bibliografskih podataka u JSON formatu. Potrebno je dogovoriti i da li će se i koji koristiti kontrolirani rječnici, koliko će biti temeljnih polja, a koliko će biti otvoreno za dodavanje. Postoji još sigurno detalja koji bi se kroz praksu i korištenjem uočili i elemenata koji bi se trebali nadograditi.

U svakom slučaju, knjižničari bi trebali sudjelovati u suvremenoj revoluciji organizacije informacija i upravljanja podacima. Kao informacijski stručnjaci trebali bi biti u centru osmišljavanja formata za zapisivanje informacija koji bi podupirali globalne standarde u relevantnim domenama. Naposljetku, ukoliko to ne učini knjižničarska struka i znanost, to će učiniti tvrtke poput Googlea ili Amazona i to vjerojatnije na manje kvalitetan način.

8 Literatura

1. Adams, Cameron. AJAX: Usable Interactivity with Remote Scripting. 2005. URL: https://www.sitepoint.com/remote-scripting-ajax/
2. Aghaei, Sareh, Mohammad Ali Nematbakhsh, and Hadi Khosravi Farsani. Evolution of the world wide web: From web 1.0 to web 4.0. // International Journal of Web & Semantic Technology 3, 1(2012.). URL: http://www.ftsm.ukm.my/ss/Book/EVOLUTION%20OF%20WWW.pdf
3. Andersen, Per. What is Web 2.0? Ideas, technologies and implications for education. // ISC Technology and Standards Watch 1, 1(2007.).
4. Anderson, Tim. Introducing XML. 2004. URL: http://www.itwriting.com/xmlintro.php
5. Architecture of the World Wide Web, Volume One. // W3C Recommendation. 2004. URL: https://www.w3.org/TR/webarch/
6. Berners-Lee, Tim. Hendler, James. Lassila, Ora. The Semantic Web. // Scientific American. Scientific American, a Division of Nature America, 2001. URL: https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf
7. Bell, Suzanne S. Librarian's Guide to Online Searching: Cultivating Database Skills for Research and Instruction. 4. Izd. Santa Barbara; Denver; Oxford: Libraries Unlimited, 2015.
8. Best, David. Web 2.0: Next big thing or next big internet bubble. // Technical report, Technische Universiteit Eindhoven, 2006.
9. Bianchini, Carlo; Guerrini, Mauro. From Bibliographic Models to Cataloging Rules: Remarks on FRBR, ICP, ISBD, and RDA and the Relationships Between Them. // Cataloging & Classification Quarterly 47, 2(2009.). URL: http://eprints.rclis.org/16303/1/CCQ_Bianchini-Guerrini_From_907343194-1.pdf
10. Bibliography. Encyclopædia Britannica. URL: https://www.britannica.com/topic/bibliography
11. Bizer, Christian, Tom Heath, and Tim Berners-Lee. Linked data: the story so far. // Semantic services, interoperability and web applications: Emerging concepts. IGI Global, 2011.

12. Clarke, Rachel Ivy. Breaking Records: The History of Bibliographic Records and Their Influence in Conceptualizing Bibliographic Data // Cataloging & Classification Quarterly 53, 3-4 (2014.).
13. Choudhury, Nupur. World wide web and its journey from web 1.0 to web 4.0. // Internationaln Journal of Computer Science and Engineering 5, 6(2014.). URL: http://ijcsit.com/docs/Volume%205/vol5issue06/ijcsit20140506265.pdf
14. Christophides, Vassilis, Vasilis Efthymiou, and Kostas Stefanidis. Entity resolution in the web of data. // Synthesis Lectures on the Semantic Web 5, 3(2015.).
15. Doan, AnHai. Halevy, Alon. Ives, Zachary. Principles of Data Integration. Waltham, MA: Morgan Kaufmann, 2012. URL: https://books.google.hr/books?id=s2YCKGrO10YC&printsec=frontcover&redir_esc=y#v=onepage&q&f=false
16. Feather, John; Sturges, Paul, eds. International Encyclopedia of Information and Library Science. 2. izd. London: Routledge. 2003.
17. Gruber, Tom. Collective knowledge systems: Where the social web meets the semantic web. // Web semantics: science, services and agents on the World Wide Web 6,.1 (2008.).
18. Gutierrez, Claudio. Modeling the Web of Data (introductory overview). // Reasoning Web. Semantic Technologies for the Web of Data. Springer Berlin Heidelberg, 2011. URL: http://citeseerx.ist.psu.edu/viewdoc/download?rep=rep1&type=pdf&doi=10.1.1.221.7285
19. Gonzales, Brighid M. Linking libraries to the web: linked data and the future of the bibliographic record. // Information Technology and Libraries (Online) 33, 4 (2014). URL: https://ejournals.bc.edu/ojs/index.php/ital/article/download/5631/pdf
20. Harris, Michael H. History of libraries in the western world. 4th ed. Metuchen: Scarecrow Press, 1995.
21. Horvat, Aleksandra. Knjižnični katalog i autorstvo. Rijeka: Naklada Benja, 1995.
22. How to do BibJSON. URL: http://okfnlabs.org/bibjson/#purpose
23. History of the Web. // World Wide Web foundation. URL: https://webfoundation.org/about/vision/history-of-the-web/
24. Interoperabilnost. // Hrvatska enciklopedija. URL: http://www.enciklopedija.hr/natuknica.aspx?id=27658

25. Introduction to XML. // w3schools.com. URL: https://www.w3schools.com/xml/xml_whatis.asp
26. Introducing JSON. // Json.org. URL: http://json.org/
27. ISBD: međunarodni standardni bibliografski opis / preporučila Skupina za pregled ISBD-a; odobrio Stalni odbor IFLA-ine Sekcije za katalogizaciju; [s engleskog prevela i predgovor napisala Ana Barbarić]. Objedinjeno izd. Zagreb: Hrvatsko knjižničarsko društvo, 2014.
28. JSON Syntax. // w3schools.com. URL: https://www.w3schools.com/js/js_json_syntax.asp
29. JSON-LD 1.0. // W3C Recommendation. 2014. URL: https://www.w3.org/TR/json-ld/#introduction
30. Katić, Tinka. Stara knjiga: bibliografska organizacija informacija. Zagreb: Hrvatsko knjižničarsko društvo, 2007.
31. Kubicek, Herbert, Ralf Cimander, and Hans Jochen Scholl. Organizational interoperability in e-government: lessons from 77 European good-practice cases. Springer Science & Business Media, 2011. URL: https://www.springer.com/cda/content/document/cda_downloaddocument/9783642225017-c7.pdf?SGWID=0-0-45-1229048-p174132406
32. Kuck, G. Tim Berners-Lee's Semantic Web. // SA Journal of Information Management 6, 1(2004.). URL: https://sajim.co.za/index.php/sajim/article/download/297/288
33. Kusserow, Arne; Groppe, Sven. Getting Indexed by Bibliographic Databases in the Area of Computer Science. // Open Journal of Web Technologies (OJWT)1, 2 (2014.).
34. Le Boeuf, Patrick. FRBR: Hype or cure-all? Introduction. // *Cataloging & Classification Quarterly* 39, 3-4 (2005.).
35. Lewandowski, Dirk. Using search engine technology to improve library catalogs. // Advances in Librarianship. Emerald Group Publishing Limited, 2010. URL: https://arxiv.org/pdf/1511.05808.pdf
36. MARC XML // The Library of Congess. URL: https://www.loc.gov/standards/marcxml/marcxml-overview.html
37. Markup Language. // The Tech Terms Computer Dictionary. URL: https://techterms.com/definition/markup_language

38. Mason, Moyak. Publishers Catalogues and Collection Development. URL: http://www.moyak.com/papers/collection-development-publishers.html
39. MODS: Uses and Features. // The Library of Congess. URL: https://www.loc.gov/standards/mods/mods-overview.html
40. Mohn, Elizabeth. Semantic Web. Salem Press Encyclopedia of Science, (2015.). URL: http://eds.b.ebscohost.com/eds/detail/detail?vid=2&sid=d6b0995c-6019-4b08-bed5-5f2c3c77870b%40sessionmgr120&bdata=Jmxhbmc9aHImc2l0ZT1lZHMtbGl2ZQ%3d%3d#AN=87323271&db=ers
41. Mohn, Elizabeth. XML (Extensible Markup Language). // *Salem Press Encyclopedia of Science*, 2015. URL: http://eds.b.ebscohost.com/eds/detail/detail?vid=2&sid=6a1742de-2f77-4fdd-aa22-ac20695d15f5%40sessionmgr102&bdata=Jmxhbmc9aHImc2l0ZT1lZHMtbGl2ZQ%3d%3d#AN=89409674&db=ers
42. Mueller, John. Understanding SOAP and REST Basics and Differences. SmartBear. 2013. URL: https://blog.smartbear.com/apis/understanding-soap-and-rest-basics/ (2018-05-16)
43. Nurseitov, Nurzhan. Paulson, Michael. Reynolds, Randall. Izurieta, Clemente. Comparison of JSON and XML Data Interchange Formats: A Case Study. // *Caine* 9, (2009.). URL: https://www.cs.montana.edu/izurieta/pubs/caine2009.pdf
44. O'Reilly, Tim. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. // International Journal of Digital Economics 65, (2007.). URL: https://mpra.ub.uni-muenchen.de/4578/1/mpra_paper_4578.pdf
45. OPAC. Hrvatska enciklopedija URL: http://www.enciklopedija.hr/natuknica.aspx?id=45194
46. Peng, Dunlu, Lidong Cao, and Wenjie Xu. Using JSON for data exchanging in web service applications. // Journal of Computational Information Systems 7, 16 (2011.). URL: https://www.researchgate.net/publication/265874991_Using_JSON_for_Data_Exchanging_in_Web_Service_Applications
47. Podatak.// Hrvatska enciklopedija. URL: http://enciklopedija.hr/Natuknica.aspx?ID=48887

48. Reitz, Joan M. Bibliographic database. // Dictionary for Library and Information Science. Westport, Connecticut: Libraries Unlimited, 2004.
49. Syntax. // Webopedia. URL: https://www.webopedia.com/TERM/S/syntax.html
50. Taylor, Arlene G; Jourdey, Daniel N. The Organization of Information. 3rd ed. Westport, Connecticut: Libraries Unlimited, 2009.
51. The History and Purpose of MARC. URL: http://lili.org/forlibs/ce/able/course8/03purpose.htm
52. UNIMARC: bibliografski format / [autorizirani prijevod IFLA [i. e.] International Federation of Library Associatons and Institutions ; hrvatski prijevod formata iz 1999. osuvremenila Marijana Tomić; stručna redakcija prijevoda Mirna Willer]. Zagreb: Hrvatsko knjižničarsko društvo ; Zadar : Sveučilište, 2009.
53. Willer, Mirna. UNIMARC u teoriji i praksi. Rijeka: Naklada Benja, 1996.
54. Xie, I., Matiusak, K. Discover digital libraries: Theory and practice. Elsevier, 2016.
55. XML Linking Language (XLink) Version 1.1. // W3C Recommendation. 2010. URL: https://www.w3.org/TR/xlink/#concepts
56. XSD (XML Schema Definition). // SearchMicroservices. URL: https://searchmicroservices.techtarget.com/definition/XSD-XML-Schema-Definition
57. Zhang, Cui-xiao, et al. Data Exchange based on Web Services. // International Journal of Computer Science and Network Security 6, 5A (2006.).

Interoperability and exchange of bibliographic data

Abstract

The thesis explores the possibilities of interoperability, exchange and reuse of bibliographic data by using current interchange formats. Aim of this work is to explain the importance of data interoperability with focus on bibliographic data. The analysis of data exchange formats is focused on use of XML and JSON as well as formats for writing and sharing bibliographic data in the context of current information services. In addition, the ability to combine these formats with traditional library standard such as representing MARC in MARCXML will be described and analyzed. Today, it is important to generate different forms of structured data for dissemination and management purposes to enable different usage scenarios. Therefore, the thesis presents the advantages of contemporary formats for sharing and writing bibliographic data in different contexts of their use. The thesis will also present some of the problems of conversion of bibliographic data record from MARC to some other contemporary forms such as the problem of complex and nested values or the occurrence of structural punctuation in the values (which often come from ISBD rules). Key terms such as data exchange formats (XML, JSON, CSV), library standards in these formats (MARCXML, MODS), Application Program Interfaces (APIs), standards, data interoperability and other related concepts will also be defined.

Keywords: bibliographic data, data exchange, interoperability, data exchange formats.