



MATEMATIČKI FAKULTET
UNIVERZITETA U BEOGRADU

DIPLOMSKI MASTER RAD

**Primena XML tehnologija u
digitalizaciji nacionalne
kulture baštine**

Autor:
Bojan
MARINKOVIĆ

Mentor:
Dr. Gordana
PAVLOVIĆ-LAŽETIĆ

Septembar, 2008

Sadržaj

1	Uvod	5
2	Preporuka formata meta-podataka	7
2.1	Trenutna situacija u Srbiji	7
2.2	Pregled međunarodnih standardih	7
2.3	Metodologija	9
2.4	Opis preporuke formata meta-podataka za opis pokretnih kulturnih dobara	9
2.5	Mapiranje u međunarodne standarde	11
3	XML shema	13
3.1	Osnove XML sheme	13
3.2	Alati za razvoj XML sheme	14
3.2.1	<oXygen/> XML Editor	14
3.2.2	SYBASE PowerDesigner 12.5	15
3.3	Reprezentacija i prezentacija meta-podataka	15
4	Pilot projekat	25
4.1	Platforma	25
4.1.1	PHP	25
4.1.2	eXist	25
4.1.3	PheXist - paket za komunikaciju sa eXist bazom	26
4.2	Realizacija	27
5	Zaključci i dalji rad	33
A	Primer	37

Glava 1

Uvod

U velikom broju institucija u Srbiji čuvaju se vrednosti iz nacionalne naučne i kulturne baštine od kojih su mnoge od velikog nacionalnog i međunarodnog značaja. Prilikom klasičnog korišćenja ovih dobara nailazimo na različite probleme, koje za posledicu imaju činjenicu da mnoga od njih ostaju poznata samo uskom krugu ljudi. Savremena tehnologija omogućava nam efikasno organizovanje sistema informacija za kulturna i naučna dobra, čime se pored dostupnosti široj publici, stručnoj i naučnoj javnosti daju i nove mogućnosti u čuvanju i zaštiti ovih dobara. Ključno mesto u ovoj oblasti ima digitalizacija - proces kreiranja digitalnih kopija sa pratećim standardom meta-podataka i softverom za manipulisanje tako definisanim meta-podacima.

Dokumentom „*Recommendations for coordination of digitization of cultural heritage in South-Eastern Europe*” [14], prihvaćenom na *South-Eastern Europe Regional meeting on digitization of cultural heritage* (Ohrid, Makedonija, 17-20. mart 2005.), konstatovano je da trenutna praksa u oblasti digitalizacije u jugoistočnoj Evropi nije na nivou EU i da bogato kulturno nasleđe nije dovoljno prezentovano u elektronskom obliku. U jednom od glavnih principa prihvaćenih na ovom sastanku se navodi da je „Poznavanje kulturne i naučne baštine važno prilikom donošenja odluka o njenoj digitalizaciji, kao i prilikom proučavanja elektronskog sadržaja. Iz ovih razloga, pravljenje popisa i kataloga bi trebalo da prethodi ili prati samu digitalizaciju kulturne i naučne baštine”.

U skladu sa ovim zaključcima Nacionalni centar za digitalizaciju (NCD, [10]) istakao je problem meta-podataka kao najvažniji u procesu izrade kataloga.

Postoji veliki broj šema meta-podataka za opis digitalizovane kulturne baštine. Prema prirodi nastanka i čuvanja, kulturna i naučna dobra u Srbiji su opisana u skladu sa različitim standardima. Pre uključivanja u nacionalnu bazu podataka, kulturna i naučna dobra je potrebno opisati jedin-

stvenim standardom. Ovaj standard bi trebalo da garantuje interoperabilnost između podataka koji potiču iz različitih izvora i kompatibilnost sa najčešće korišćenim međunarodnim standardima. Zbog svega rečenog, predlog standarda se sastoji iz sledeća dva dela:

- opisa meta-podataka koji se odnose na pojedinačna pokretna kulturna dobra i
- šeme prevođenja u/iz nekih najčešće korišćenih međunarodnih standarda.

Nacionalni komitet UNESCO-a za digitalizaciju prepoznao je značaj ovog predloga i prihvatio ga kao preporuku formata meta-podataka na zajedničkom sastanku NCD-a i Nacionalnog komiteta UNESCO-a za digitalizaciju, održanom 1. marta 2007. godine.

Pre pravljenja centralnog nacionalnog kataloga kulturnih dobara, odlučeno je da se započne pilot projekat kojim bi se ispitale mogućnosti primene XML (eXtensible Markup Language) tehnologija u oblasti digitalizacije. Kako se može videti i u [18], XML tehnologije su logičan izbor za predstavljanje meta-podataka u mašinski čitljivoj formi, jer je obrada ovako prikupljenih podataka značajno pojednostavljena.

U ovom radu biće prikazani predlog formata meta-podataka za opis pokretnih kulturnih dobara, metodologija njegove izrade, XML shema nastala na osnovu ovog formata kao i aplikacija nastala kao rezultat pilot projekta za pravljenje nacionalnog kataloga pokretnih kulturnih dobara. Sve navedeno je izrađeno od strane tima sastavljenog od predstavnika svih institucija koje su formirale NCD. Autor ovog rada, kao član pomenutog tima, bio je zadužen za izradu XML sheme zasnove na predlogu formata meta-podataka i realizaciju aplikacije pilot projekta.

Glava 2

Preporuka formata meta-podataka za opis pokretnih kulturnih dobara

2.1 Trenutna situacija u Srbiji

Trenutno u Srbiji ne postoji standard meta-podataka vezan za digitalizovana kulturna dobra koji je u širokoj upotrebi. Takođe, iako je proces digitalizacije započet u izvesnom broju institucija koje se staraju o kulturnoj baštini, trenutno ne postoji standard meta-podataka koji je formalno prihvaćen na državnom nivou. Različiti izvori kulturnih i naučnih dobara (biblioteke, muzeji, arhivi, neki istraživački centri) koriste međunarodne standarde koji su odgovarajući u njihovim oblastima, ad-hoc metode ili zastarele procedure za opis kulturnih dobara u klasičnom obliku (definisane tokom 80-tih i početkom 90-tih godina prošlog veka). Neke institucije, pak, čekaju rešenje ovog problema vezanog za meta-podatke i ne preduzimaju nikakve korake na polju digitalizacije. Ovo sve za posledicu ima da se, ukoliko postoje, digitalni katalozi u Srbiji ne mogu koristiti među institucijama različitih tipova.

2.2 Pregled međunarodnih standardih

Postoji veliki broj međunarodnih standarda za opis kulturnih dobara, kao, na primer:

- Dublin Core [5]
- EAD [6]

- MARC [9]
- TEL AP [13]
- CIDOC CRM [7]
- FRBR [3, 8], itd.

Dublin Core razvija *Dublin Core Metadata Initiative* (DCMI). Cilj DCMI je da promoviše interoperabilnost standarda meta-podataka i razvoj specijalizovanog rečnika meta-podataka za opis resursa. Iako je ovaj standard moguće primeniti na sve tipove dobara osiguravajući interoperabilnost, kako je prikazano u [4], problem je što je DC skup podataka (*Simple DC*) prilično restriktivan, pa se različite informacije moraju grupisati unutar jednog elementa. Nedavno je iz DCMI predstavljeno poboljšanje *Simple DC*-a, nazvano *Qualified DC*, takvo da je moguće dodavati nove elemente u *Simple DC* koji su specifični za pojedine aplikacije i domene. Ono što je potrebno naglasiti je da je moguće izgubiti pojedine informacije prilikom procesa prevođenja vrednosti iz *Qualified DC* u *Simple DC*.

Ostali pomenuti standardi uglavnom su vezani za pojedinačne tipove kulturnih dobara. *Encoded Archival Description* (EAD) je formulisan za arhivske dokumente sa ciljem da ih približi korisnicima. Slično, MARC standard se koristi za prezentovanje bibliografskih i sličnih informacija u mašinski čitljivom obliku. Još jedan bibliotečki standard je *The European Library Application Profile* (TEL AP). TEL AP je baziran na predlogu DCMI radne grupe za bibliotekarstvo iz 2002. Narodna biblioteka Srbije je od 2005. punopravni član TEL grupe. Trebalo bi naglasiti da je TEL AP još uvek u fazi razvoja.

Functional requirements for bibliographic records (FRBR) je konceptualni, objektno-orijentisani model meta-podataka, takođe nastao u svetu biblioteka. Model se sastoji iz tri grupe entiteta. Prvu grupu čine produkti intelektualnog ili umetničkog nastanka koji se opisuju: rad, izraz, manifestacija i stvar. Druga grupa, koju čine osobe i udruženja, povezana je sa intelektualnim ili umetničkim sadržajem, proizvodnjom i distribucijom, kao i nadzorom nad ovim proizvodima. Treća grupa obuhvata koncept, objekat, događaj i mesto. FRBR je proširiv na arhive, muzeje i oblast izdavaštva.

U oblasti muzeja ne postoji široko prihvaćeni standard meta-podataka. Postoji skup preporuka nazvan *Spectrum*, britanski standard o muzejskoj građi [12]. Iako su ove preporuke veoma važne i korisne za odgovarajući poddomen kulturnog blaga, one se ne mogu direktno primeniti na ceo domen.

CIDOC object-oriented Conceptual Reference Model (CIDOC CRM), razvijen je od strane *ICOM/CIDOC Documentation Standards Group*. Zajedničkim radom *CIDOC CRM SIG* i *ISO/TC46/SC4/WG9*, CIDOC CRM

je proglašen ISO standardom septembra 2000. Septembra 2006. prihvaćen je kao ISO 21127. CIDOC CRM predstavlja ontologiju za informacije o kulturnim dobrima, tj. formalnim jezikom opisuje, kako eksplicitne, tako i implicitne koncepte i relacije relevantne za dokumentaciju o kulturnim dobrima. Primarna uloga CIDOC CRM je da posluži kao osnova za uređenje informacija o kulturnim dobrima, kao i da pruži semantiku potrebnu za transformaciju dosadašnjih lokalizovanih izvora u jednu koherentnu globalnu celinu.

Na osnovu svega ovoga, NCD je odlučio da definiše jedinstveni nacionalni standard meta-podataka za opis pokretne kulturne baštine, kako bi se osigurala interoperabilnost između različitih digitalnih resursa nacionalne kulturne i naučne baštine.

2.3 Metodologija

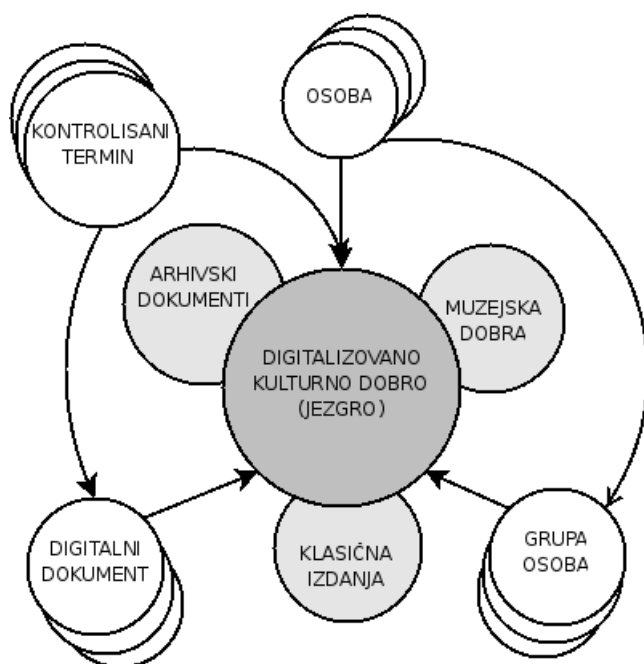
Tokom razvoja standarda postavljeni su uslovi kako bi se prevazišle razlike prilikom opisivanja, kao i da bi se obezbedilo obuvatanje specifičnih potreba različitih izvora. Ti uslovi su da:

- definicija mora biti dovoljno bogata kako bi se opisali sadržaji biblioteka, muzeja i arhiva, kao i drugih (na primer, istraživačkih) ustanova
- definicija mora biti dovoljno fleksibilna da bi se obezbedio prevod u/iz međunarodnih standarda
- definicija mora da omogući: višejezični opis kulturnih i naučnih dobara, kao i korišćenje predefinisanih rečenika za pojedinačne elemente kada god je to moguće.

2.4 Opis preporuke formata meta-podataka za opis pokretnih kulturnih dobara

Osnovni tipovi objekta objektnog modela su:

- Osoba i Grupa osoba – koristi se za predstavljanje entiteta koji imaju intelektualni ili na drugi način važan doprinos u nastanku kulturnog dobra ili su vlasnici kulturnog dobra
- Digitalni dokument – za predstavljanje pojedinačnog dokumenta koji sadrži digitalnu reprezentaciju nekog kulturnog dobra
- Digitalizovano kulturno dobro



Slika 2.1: Odnos između različitih elemenata standarda

- Kontrolisani termin – za predstavljanje elemenata predefinisano rečnika

Slika 2.1 prikazuje strukturu objektnog modela meta-podataka, u okviru kojeg čvorovi predstavljaju tipove objekata, dok su lukovima predstavljene mogućnosti preuzimanja pojedinih podataka između objekata različitih tipova. Ova struktura formira jedinstveno jezgro reprezentujući presek opisa dobara iz biblioteka, arhiva i muzeja. Takođe, ovaj predlog dozvoljava i neka proširenja vezana za pojedinačna polja od interesa. Objekti tipa Kontrolisani termin odgovaraju standardizovanom skupu vrednosti koje poboljšavaju pristup informacijama o kulturnom dobru. Ove vrednosti bi trebalo organizovati u strukturno kontrolisani rečnik koji nudi dozvoljene termine i sinonime, kao, na primer, u [2]. Definisani su skupovi elemenata koji odgovaraju objektima. Oni se sastoje iz opisnog i administrativnog dela. Administrativni deo identifikuje autore i/ili vlasnike opisa, svojstva kulturnog dobra i odgovarajuće digitalne reprezentacije. Treba napomenuti da, ukoliko je potrebno, moguće je opisati i prava pristupa kako bi se ograničio pristup izvoru. Kompletan verzija opisa preporuke formata meta-podataka za opis pokretnih kulturnih dobara se nalazi na adresi [11]. Primer opisa jednog kulturnog dobra prema ovoj preporuci dat je dodatku A ovog rada.

NCD	DC	ISAD	EAD	TEL AP	Library of Congress
Description	description	3.3.1	scopecontent	description	description summary
Type	type	3.1.5	genreform	type	original content type
Material	medium	3.1.5	physdesc	format. medium	
Rights	rights	3.4.2	userrestrict	rights	access rights
Access rights	accessRights	3.4.1	accessrestrict	rights	access category
Note		3.6.1	note, odd		
Capture device(s)			daodesc		capture device ID
MIME format(s)	Format		daodesc	format	internet media type

Tabela 2.1: Mapiranje u/iz međunarodnih standarda

2.5 Mapiranje u međunarodne standarde

Kao poseban deo, ova preporuka formata meta-podataka definiše i prevođenje u i iz nekih međunarodnih standarda meta-podataka (*Dublin Core Metadata Element Set, The European Library Application Profile, Table of Core Metadata Elements for Library of Congress Digital Repository Development, Encoded Archival Description*). Primer jednog prevođenja dat je u tabeli 2.1. Kao što se može primetiti, ovo prevođenje nije jedinstveno pošto jednom elementu ove preporuke može odgovarati nijedan ili, pak, više odgovarajućih elemenata drugih standarda, ali i više elemenata ovog predloga može sadržati informacije koje odgovoraju samo jednom elementu nekog od navedenih standarda.

Glava 3

XML shema

3.1 Osnove XML sheme

Extensible Markup Language (XML) je specifikacija opšte namene za definisanje jezika za označavanje. Definiciju pojedinačne specifikacije moguće je dati na različite načine. XML shema je jedan od mogućih načina.

XML shema objavljena je kao preporuka W3C konzorcijuma u maju 2001. godine. To je bio prvi nezavisan shema jezik koji je u potpunosti zadovoljio sve preporuke koje je W3C postavio. Kao i svi XML jezici, XML shema se može koristiti radi izražavanja neke sheme, tj. skupa pravila koje mora zadovoljiti XML dokument da bi bio konzistentan i validan u skladu sa datom shemom. Pod konzistentnošću se podrazumeva sintaksna ispravnost nekog XML dokumenta. Da bi dokument bio i validan potrebno je da svi upotrebljeni elementi i atributi zadovoljavaju odgovorajuću strukturu, kao i da svi elementi i atributi za koje se zahteva da budu korišćeni zaista budu i upotrebljeni. Takođe, proverava se da li vrednosti strukture za koje su definisani regularni izrazi zadovoljavaju te regularne izraze, kao i tipovi i vrednosti svih upotrebljenih elementa i atributa.

Instanca XML sheme je definicija XML sheme (XSD) sa uobičajnom ekstenzijom fajla ".xsd", tako da se i sam jezik često i naziva XSD. Tokom razvoja, XML shema je bila pod uticajem definicije tipa dokumenta (engl. **D**ocument **T**ype **D**efinition, DTD). DTD je jedan od jezika iz sistema za definisanje jezika za označavanje (engl. **S**tandard **G**eneralized **M**arkup **L**anguage, SGML). Primarno korišćenje DTD-a je za izražavanje sheme kao skupa pravila, koja definišu sintaksu klasa, tipova i strukture nekog dokumenta. DTD je jezik manje izražajnosti od XML sheme. Postojali su i drugi pokušaji da se naprave XML sheme, kao što su DDML, SOX, XML-Data i XDR. Može se reći da su prilikom definisanja XML shema uzeti delovi svakog od nave-

denih jezika, kao i da je cilj bio da se prevaziđu razlike među navedenim jezicima.

Nakon osnovne XML shema validacije, moguće je izraziti strukturu XML dokumenta i sadržaj u formi modela podataka. Model podataka XML sheme uključuje[16]:

- rečnik (imena elemenata i atributa)
- model sadržaja (relacije i strukturu)
- tipove podataka.

Ovako definisana kolekcija podataka naziva se skupom podataka nakon validacije sheme (engl. **Post-Schema-Validation Infoset**, PSVI). Ono što je značajno istaći je da PSVI tretira dokument kao objekat u skladu sa objektno-orijentisanom paradigmom.

Tipovi podataka u XML shemi podeljeni su na ugrađene i one koji su definisani od strane korisnika. Postoji 19 primitivnih i 25 izvedenih ugrađenih tipova podataka. Svi ovi tipovi reprezentuju uobičajene tipove podataka, kao što su stringovi, celi i realni brojevi, kao i vremenske odrednice. Problem sa ovakvom definicijom tipova podataka XML sheme je restriktivnost, jer se svi ostali „specifični” tipovi podataka moraju izvesti iz ugrađenih, što je u nekim slučajevima veoma teško, a ponekad i nije moguće.

3.2 Alati za razvoj XML sheme

3.2.1 <oXygen/> XML Editor

<oXygen/> je zaokružena platforma koja pored editovanja XML dokumenata sadrži alate za:

- autorizaciju i prevođenje XML dokumenata,
- razvoj XML shema, DTD i Relax NG,
- debugovanje XPath, XSLT i XQuery upita,
- testiranje protokola za pristup jednostavnim objektima (engl. **Simple Object Access Protocol**, SOAP) i jezika za opis mrežnih servisa (engl. **Web Services Description Language**, WSDL).

Pomoću ovog alata moguća je integracija XML dokumenata u repozitorijume pomoću WebDAV, Subversion i S/FTP protokola. <oXygen/>

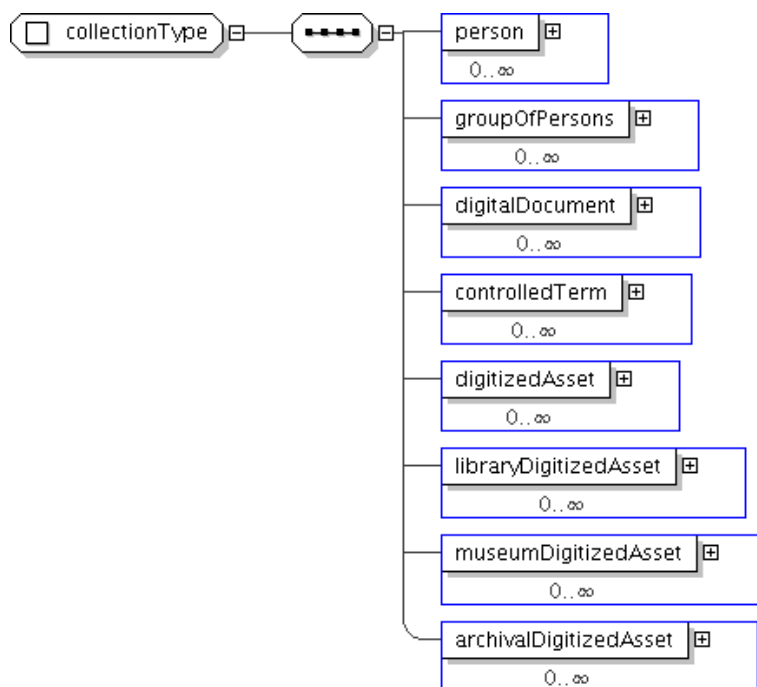
podržava i pregledanje, ažuriranje i pretraživanje XML dokumenata, ali i relacionih baza podataka. Ovaj alat dostupan je i kao Eclipse IDE plugin, što je veoma zahvalno u slučaju kreiranja Java aplikacija.

3.2.2 SYBASE PowerDesigner 12.5

PowerDesigner 12.5 je alat za razvoj modela i softvera. To je snažan softverski alat koji pored samog razvoja softvera omogućava i analizu rešenja u različitim fazama životnog ciklusa softvera. Podržava nekoliko standardnih tehnika za modelovanje podataka, kao što su UML i modelovanje poslovnih procesa, ali i modelovanje objekata pomoću XML tehnika. Ovo razvojno okruženje podržava različite programske, ali i skript jezike za realizaciju aplikacija.

3.3 Reprezentacija i prezentacija preporuke formata meta-podataka za opis pokretnih kulturnih dobara pomoću XML-a

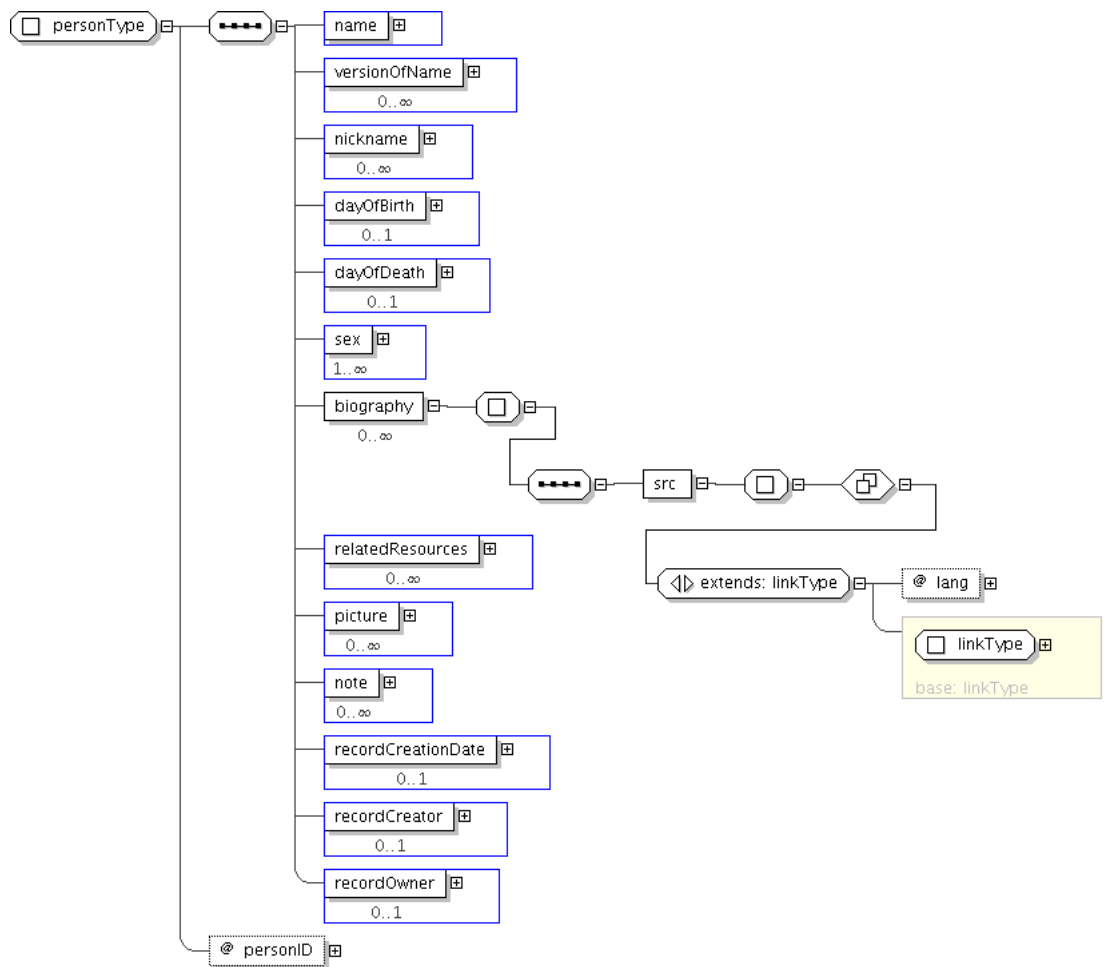
Preporukom formata meta-podataka za opis pokretnih kulturnih dobara definisano je koje bi meta-podatke trebalo navesti prilikom opisa nekog kulturnog dobra. Da bi ovi podaci bili mašinski čitljivi i dobro strukturirani, pomoću <oxygen/> XML editora i SYBASE PowerDesigner 12.5 je definisana XML shema koja odgovara ovom predlogu formata meta-podataka. SYBASE PowerDesigner 12.5 je korišćen u procesu defisanja objekta modela podataka, dok je <oxygen/> XML editor korišćen pri detaljnoj specifikaciji pojedinačnih tipova podataka.



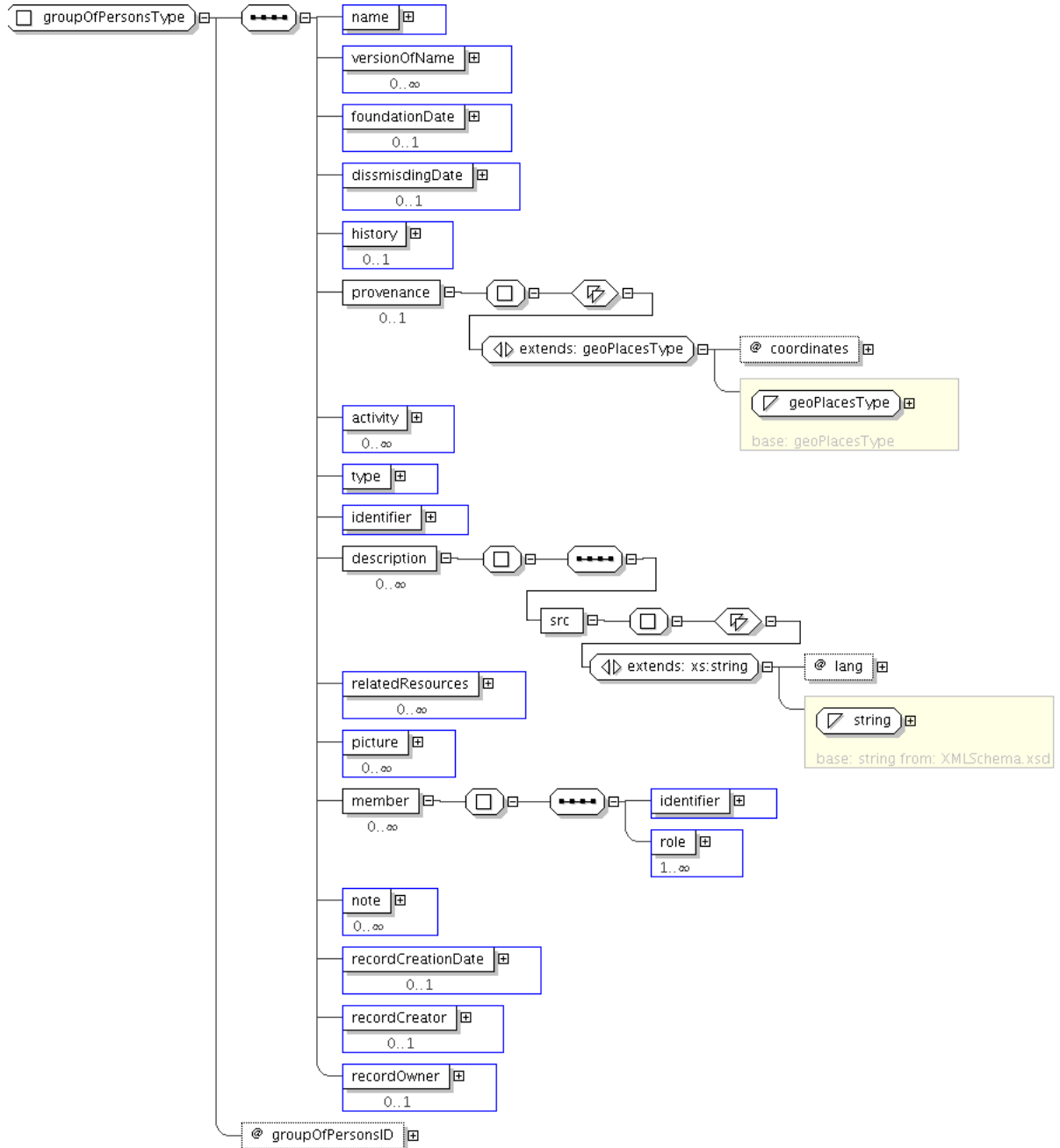
Slika 3.1: Objekat tipa - Kolekcija

Pored osnovnih tipova podataka koje zahteva standard, a koji su ilustrirani na slikama 3.1, 3.2, 3.3, 3.4, 3.5, 3.7, 3.8, 3.9 i 3.10, u samoj shemi definisano je više jednostavnih tipova podataka. Većina ovih tipova definisana je samo kao preimenovanje osnovnog `xsd:string` tipa podataka za stringove. Ovo je učinjeno zbog potrebe da se u kasnijoj fazi razvoja kataloga na višem nivou izvrši provera vrednosti, pošto pojedini elementi uzimaju vrednosti iz strogo definisanih skupova podataka. Takođe, definisano je i više kompleksnih tipova, kao što su, na primer, lična imena i datumi, koji sami za sebe mogu predstavljati zasebne celine.

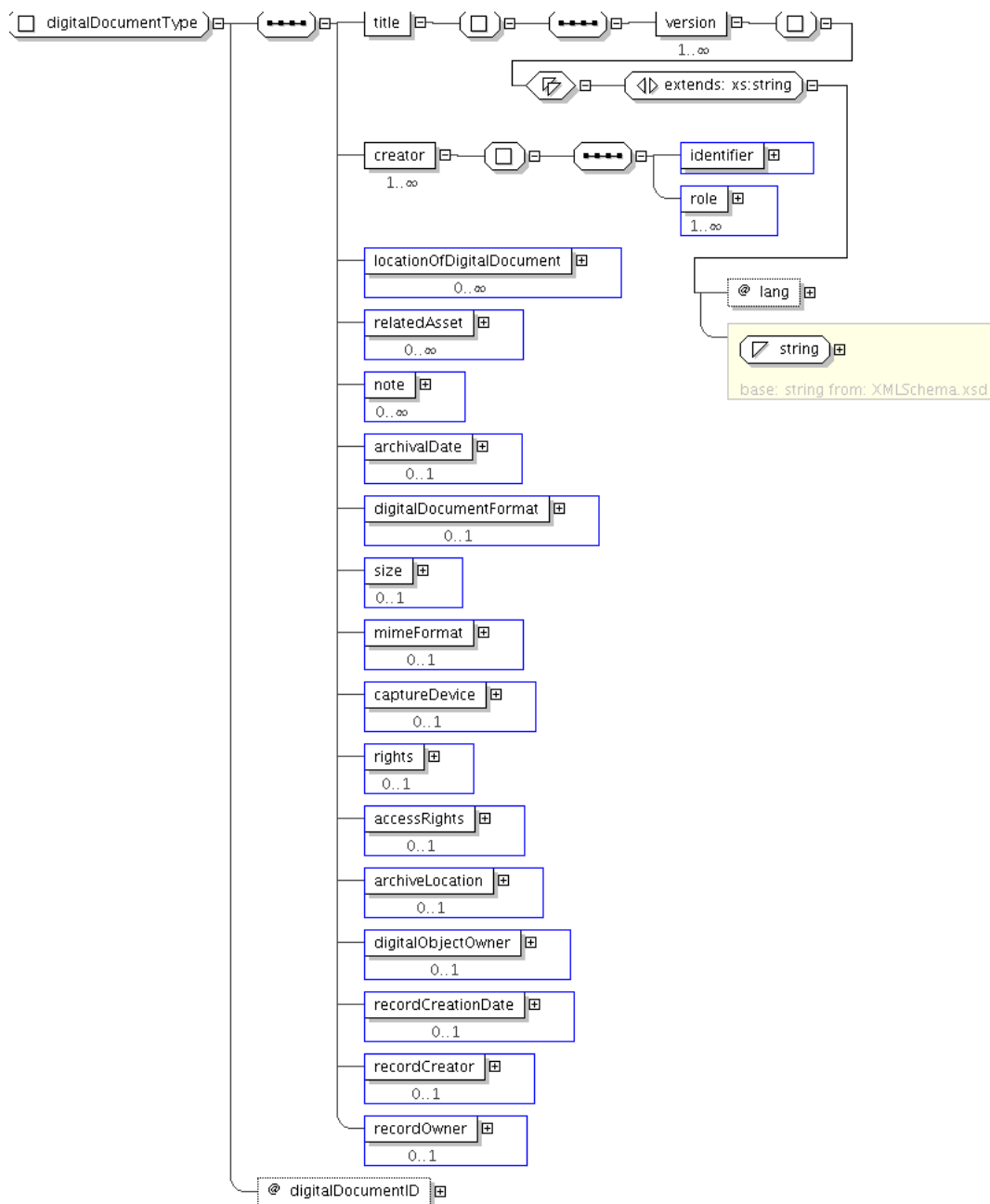
Pored ovoga, definisana je CSS datoteka. U ovoj datoteci definisana su pravila za transformaciju pojedinačnih XML tagova, kako bi podaci koji oni nose sa sobom mogli biti prezentovani u formatu koji je pogodniji za čoveka. Rezultat ove transformacije može se videti na slici 4.6.



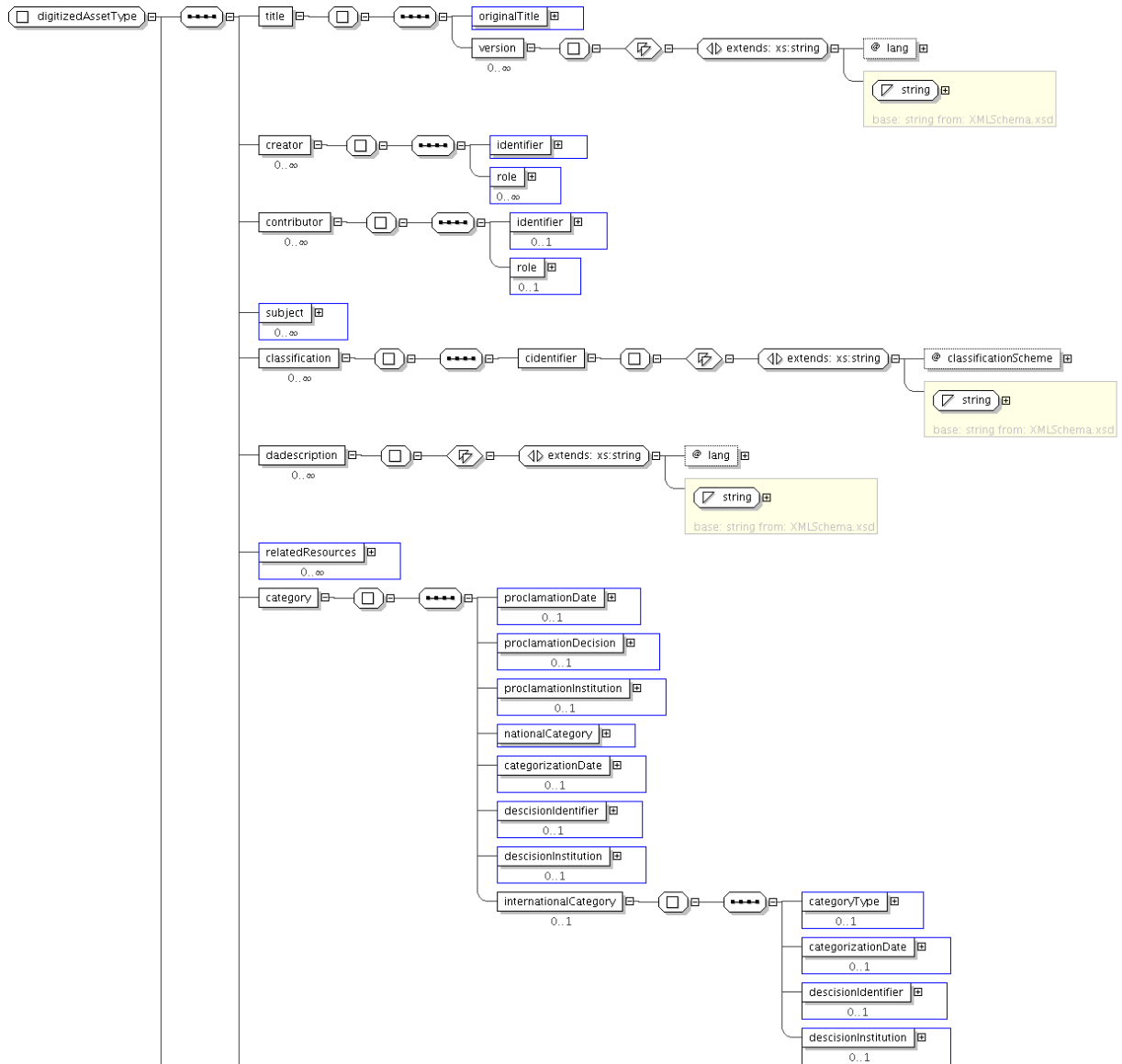
Slika 3.2: Objekat tipa - Osoba



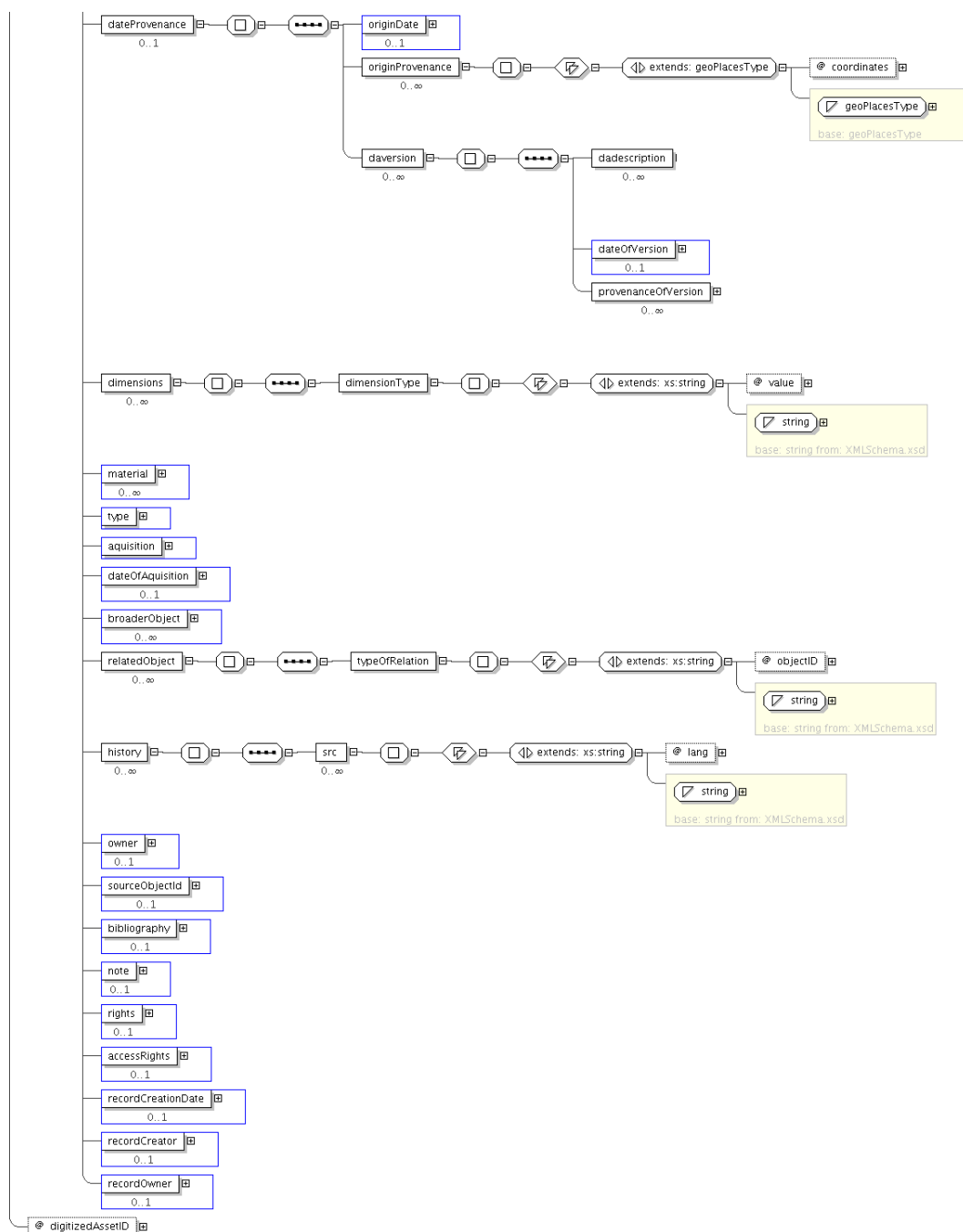
Slika 3.3: Objekat tipa - Grupa osoba



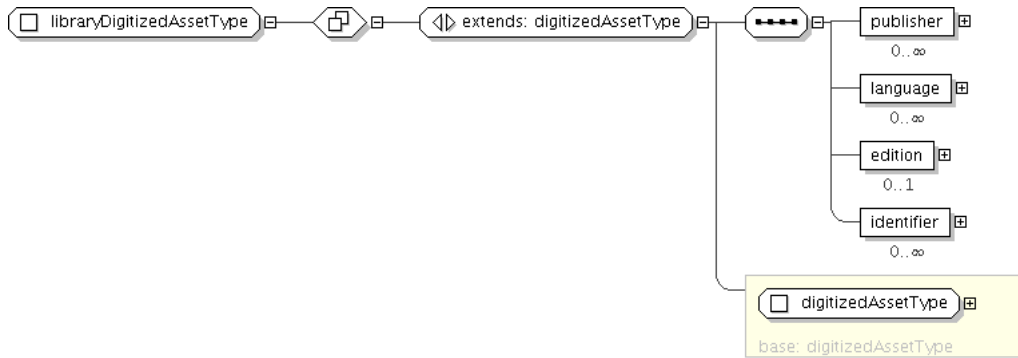
Slika 3.4: Objekat tipa - Digitalni dokument



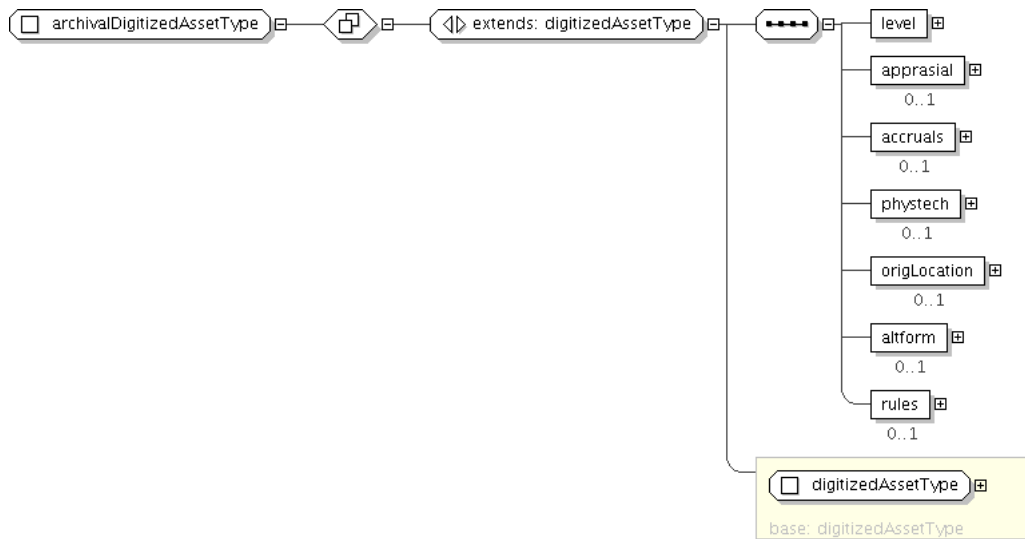
Slika 3.5: Objekat tipa - Digitalizovano kulturno dobro



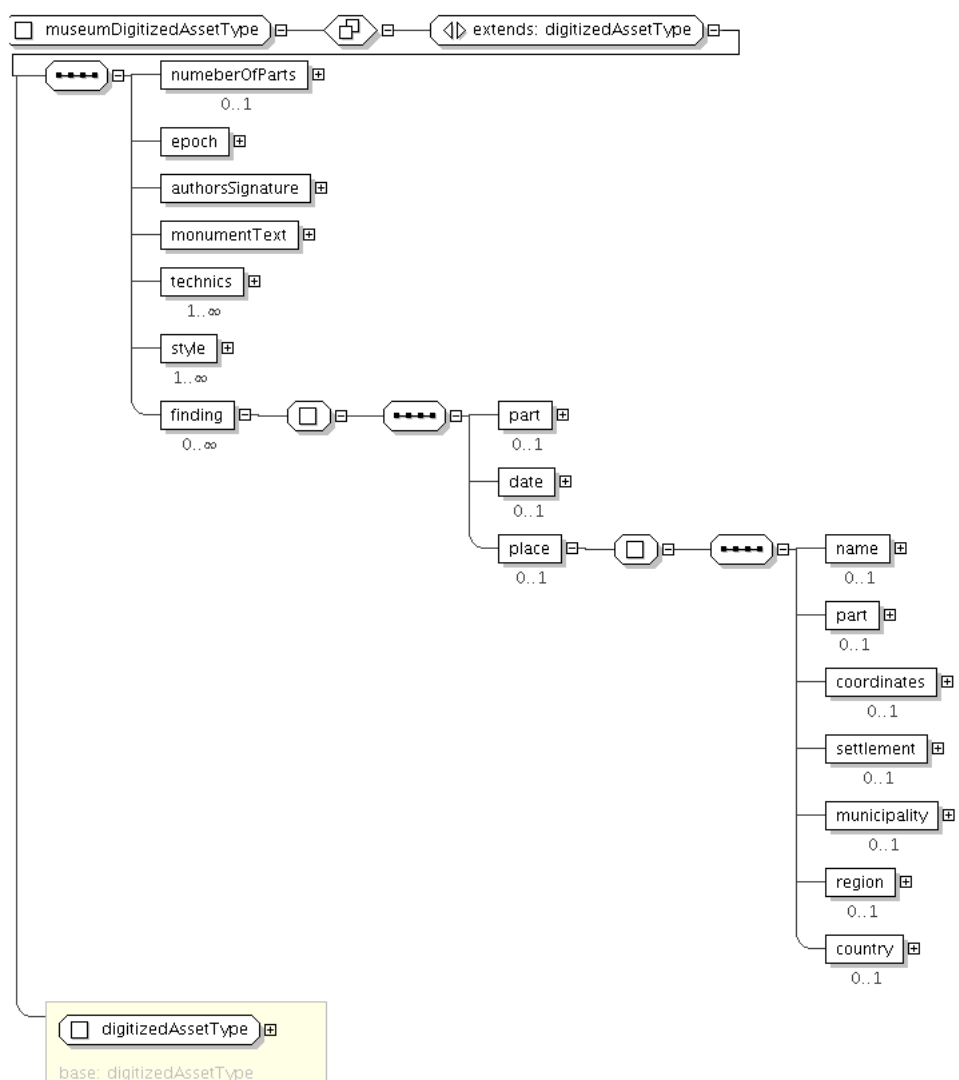
Slika 3.6: Objekat tipa - Digitalizovano kulturno dobro - nastavak



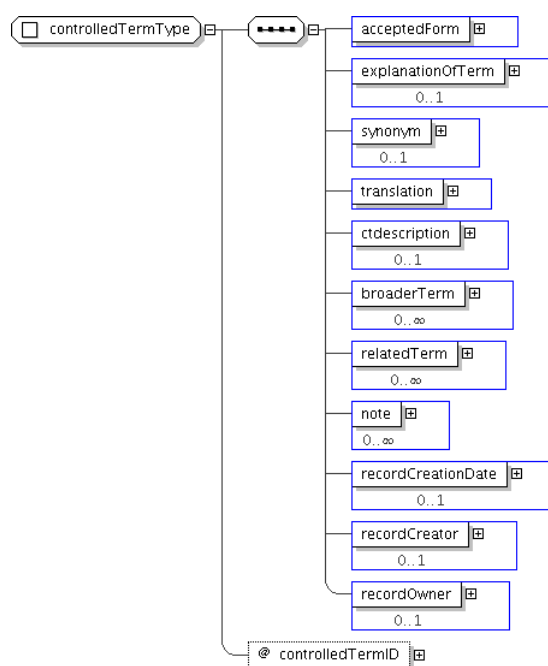
Slika 3.7: Objekat tipa - Digitalizovano kulturno dobro, klasično izdanje



Slika 3.8: Objekat tipa - Digitalizovano kulturno dobro, arhivski dokument



Slika 3.9: Objekat tipa - Digitalizovano kulturno dobro, muzejsko dobro



Slika 3.10: Objekat tipa - Kontrolisani termin

Glava 4

Pilot projekat: Katalog pokretnih kulturnih dobara

4.1 Platforma

Da bi ovaj predlog formata meta-podataka zaživeo u praksi, radna grupa NCD-a je odlučila da započne realizaciju pilot projekta pod nazivom „Katalog pokretnih kulturnih dobara”. Odlučeno je da u osnovi tog projekta bude klasična klijent-server arhitektura, nakon čega je odlučeno da se pilot projekat realizuje pomoću PHP-a, dok je za sistem za upravljanje bazama podataka izabrana izvorna XML baza podataka (engl. native XML database, NXD) eXist-db [17]. Realizacija pilot projekta prepuštena je autoru ovog rada.

4.1.1 PHP

PHP (rekurzivni akronim za *PHP: Hypertext Preprocessor*) je skript jezik, prvobitno namenjen razvoju dinamičkih internet prezentacija. Namenjen je izvršavanju na serverskoj strani, ali se može koristiti i iz komandne linije, kao i iz samostalnih GUI aplikacija. Široko je rasprostranjen kao skript jezik opšte namene koji je pogodan za razvoj u web okruženju i može biti primenjen u okviru HTML-a. Interpretacijom PHP koda serveri generišu web strane. Podržan je od strane najvećeg broja web servera i operativnih sistema, i najčešće se koristi kao modul Apache web servera.

4.1.2 eXist

eXist-db je *open source* sistem za upravljanje bazama podataka koji je u potpunosti zasnovan na XML tehnologiji. On čuva XML podatke prema

XML modelu podataka i ima razvijenu efikasnu, indeksno zasnovanu *XQuery* obradu podataka.

eXist-db podržava sledeće (web) tehnologije i standarde:

- XQuery 1.0 / XPath 2.0
- XSLT 1.0 (uz korišćenje *Apache Xalan*) i XSLT 2.0 (uz opciono korišćenje *Saxon-a*)
- HTTP interfejs: REST, WebDAV, SOAP, XMLRPC, Atom Publishing Protocol
- XML specifične baze podataka: XMLDB, XUpdate, XQuery update extensions

eXist-db je saglasan sa XQuery standardom. Pretraživač je proširiv i podržava veliku kolekciju XQuery funkcijskih modula. eXist pruža sopstvenu implementaciju XQuery-ja, koja je podržana efiksanom shemom za indeksiranje u jezgru baze, čime je omogućena brza identifikacija strukturnih relacija između čvorova.

4.1.3 PheXist - paket za komunikaciju sa eXist bazom

PheXist je skup klasa (PHP5, PHP4 & Perl) za postavljanje upita i manipulaciju XML dokumentima unutar eXist XML:DB.

Ovaj skup klasa implementiran je u PHP-u, ali i u Perl-u, čime je omogućeno postavljanje upita eXist XML:DB korišćenjem XQuery jezika.

Veza sa eXist XML:DB ostvarena je pomoću SOAP protokola.

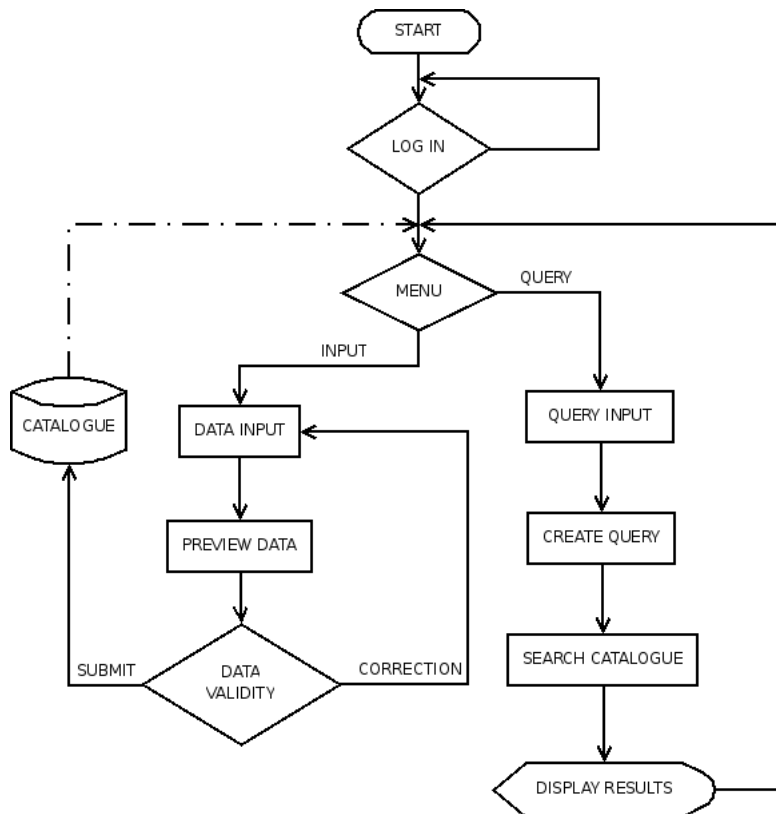
Implementirane metode su:

- Constructor(string User, string Password, string WSDL) : kreira SOAP proxy klijenta
- connect() : povezuje se sa eXist XML:DB
- disconnect() : prekida konekciju iz trenutne sesije
- xquery(string Query) : postavlja upit eXist XML:DB korišćenjem XQuery jezika
- getError() : vraća poruku o grešci (implementirana samo u PHP-u)
- debug(boolean Debug?) : vrši debugovanje komunikacije između klase i servera, kao i štampanje svih XQuery rezultata (implementirana samo u PHP-u)

- store(string XMLdata, string encoding, string path, boolean replace): unosi novi XML dokument u bazu
- removeDocument(string path): briše XML dokument iz baze
- createCollection(string path): pravi novu kolekciju XML dokumenata
- removeCollection(string path): briše kolekciju XML dokumenata

4.2 Realizacija

Kao što se može videti iz dijagrama toka programa 4.1, prvo na šta korisnik kataloga nailazi je maska za pristup (na slici 4.2).



Slika 4.1: Dijagram toka programa



Slika 4.2: Maska za pristup katalogu

U ovoj fazi projekta procenjeno je da je najbolje da se menadžment korisnika ostavi na nivou samog sistema za upravljanje bazom podataka, tako da se aplikacija obraća sistemu i proverava da li postoji korisnik baze sa unetim korisničkim imenom i lozinkom.



Slika 4.3: Maska glavnog menija

Ukoliko takav korisnik postoji, prikazuje se maska glavnog menija aplikacije prikaznog na slici 4.3. Aplikacija se sastoji iz dela za unos novih podataka u bazu i dela za pretragu već postojećih podataka.

Slika 4.4: Maska za unos novog elementa

Postoje maske za unos svakog od osnovnih objekata preporučenog formata. Na slici 4.4 može se videti deo maske za unos podataka o novoj osobi. U maskama su jasno naznačeni elementi za koje je potrebno uneti podatke, kao i elementi koji će se ponavljati u okviru nekog od osnovnih objekata, kao što je to definisano samom preporukom formata i XML shemom. Nakon unosa podataka postoji jedan među korak pre unosa samog objekta u bazu. To je provera validnosti podataka, kako od strane korisnika, tako i od strane same aplikacije.

Slika 4.5: Maska za pretragu baze podataka

Kao što se može videti na slici 4.5, pretragu sadržaja baze moguće je izvesti na dva načina:

- listanjem svih objekata koji se nalaze u bazi
- pretragom elemenata koji zadovoljavaju određeno svojstvo.

U svakom od ovih slučajeva generiše se upit u XQuery jeziku koji se šalje sistemu za upravljanje bazom podataka, koji dalje vraća rezultat pretrage aplikaciji, nakog čega se vrši transformacija dobijenih podataka iz XML formata u čitljiviju formu, kao što se može videti na slici 4.6.

MAIN MENU | NEW QUERY

Found 7 hits in 13 ms.

Result(s):

```

Person
Name
First name: Branislav
Family name: Nušić
Version of name
First name: Ben
Family name: Akiba
Day of birth: 1864-10-08
Day of death: 1938-01-19
Sex:
1
Biography
in sr:http://sr.wikipedia.org/wiki/Бранислав_Нушић
Biography
in en:http://en.wikipedia.org/wiki/Branislav_Nušić
Record creation date: 2007-03-30
Record creator: Tamara Butigan
Record owner: NBS
  
```

```

Person
Name
First name: Dragoslav
Family name: Živković
Nickname: Bob
Sex:
1
Record creation date: 2007-03-30
  
```

Slika 4.6: Prikaz rezultata upita

Trenutan izgled maske za generisanje upita omogućava pretragu po jednostavnoj selekciji. Tako se upitom:

```

for $sth in
//person | //groupOfPersons | //digitalDocument |
//digitizedAsset | //libraryDigitizedAsset |
//museumDigitizedAsset | //archivalDigitizedAsset
return $sth
  
```

kao rezultat dobijamo listu svih objekata koji se nalaze u bazi.

Upitom:

```
for $sth in
//person[ name/firstName = "Branislav" ]
return $sth
```

kao rezultat dobijamo podatke o svim osobama koje se nalaze u bazi, a čije je ime Branislav. U slučaju kada je sadržaj baze zasnovan na primeru iz dodatka A, rezultat će biti podaci koji se mogu videti u tabeli A.1.

Upitom:

```
for $sth in
//person[ recordCreator = "Tamara Butigan"]
return $sth
```

kao rezultat dobijamo podatke o svim osobama koje se nalaze u bazi, a za koje je podatke kreirala Tamara Butigan. U slučaju kada je sadržaj baze zasnovan na primeru iz dodatka A, rezultat će biti podaci koji se mogu videti u tabelama A.1, A.2, A.3 i A.4.

Maska za generisanje upita biće menjana u skladu sa budućim potrebama korisnika, kako bi oni na jednostvniji način realizovali složenije upite.

Glava 5

Zaključci i dalji rad

Glavni cilj nacionalnog standarda je da garantuje:

- interoperabilnost između resursa dostupnih iz različitih izvora
- kompatibilnost sa najpopularnijim postojećim međunarodnim standardima.

Ova preporuka formata meta-podataka je napravljena i biće ubuduće poboljšavana od strane NCD-a. Takođe, predložena je Ministarstvu kulture radi zvaničnog prihvatanja, kao format meta-podataka budućeg centralnog nacionalnog kataloga pokretnih kulturnih dobara. Očekivana zvanična podrška će, posebno, biti važna kada je u pitanju pravljenje odogovarajućeg tezaurusa.

Prilikom definisanja ove preporuke formata meta-podataka korišćen je objektno-orijentisani pristup. Koristi koje su na ovaj način ostvarene na tehničkom planu su sledeće:

- modularnost: olakšava se izmena pojedinačnih komponenti opisa
- ponovno korišćenje objekata: omogućava se višestruko korišćenje nekog objekta u sistemu (na primer, neka osoba može biti istovremeno autor nekih dobara, član neke grupe, vlasnik nekih dokumenata, itd.)

Jedan od prvih narednih koraka će biti izrada odgovarajućeg opisa u formatu *Qualified DC*.

Trenutni model poštuje pravila druge FRBR grupe entiteta koja se odnosi na osobe i grupe osoba sa svim njihovim osnovnim atributima. U budućnosti će se nastojati da se i ostali segmenti FRBR modela primene, kako bi se olakšao pristup pojedinačnim segmentima digitalne kolekcije od strane korisnika.

Kako je ovaj predlog skup meta-podataka koji se odnosi na pojedinačna pokretna kulturna dobra, plan NCD-a je da definiše sličan skup meta-podataka za opis:

- nepokretnih kulturnih dobara
- kolekcija kulturnih dobara.

Pilot projekat kataloga pokretnih kulturnih dobara je pokazao da je uz malo sredstava na efikasan način moguće obraditi informacije iz različitih izvora. Treba pomenuti da je dosadašnja implemenetacija realizovana na *open source* platformi uz korišćenje licenciranih alata. U dalji razvoj centralnog nacionalnog kataloga potrebno je uložiti više resursa. Primenom navedenih alata moguće je poboljšati sam izgled aplikacije, ali je na jednostvan način moguće realizovati i *stand-alone* aplikaciju koja bi značajno doprinela u segmentu zaštite prikupljenih informacija. Pored ovoga, neophodna je i podrška zvaničnih institucija kako bi ovaj katalog zaživeo u praksi.

Na kraju bi trebalo istaći da razvoj nacionalnog standarda meta-podataka, kao i nacionalnog kataloga pokretnih kulturnih dobara ima širi uticaj od samog opisivanja i popisivanja kulturnih dobara. Sve ovo može služiti kao podsticaj pojedincima koji se bave raznim vidovima zaštite nacionalne kulturne baštine da svoje napore usmere ka pravljenju digitalnih kopija kulturnih dobara i na taj način prevaziđu razlike i rivalstva između raznih profesija na nacionalnom i međunarodnom nivou, kako bi se u budućnosti poboljšalo prisustvo kulturne i naučne baštine iz Srbije na internetu.

Literatura

- [1] Z.Ognjanović, T. Butigan-Vučaj, B. Marinković, *NCD Recommendation for the National Standard for Describing Digitized Heritage in Serbia*, Metadata and Semantics, Edited by: Sicilia, Miguel-Angel, Lytras, Miltiadis D, Springer, 2009, ISBN: 978-0-387-77744-3
- [2] The Art & Architecture Thesaurus, AAT, http://www.getty.edu/research/conducting_research/vocabularies/aat/.
- [3] A weblog about FRBR: Functional Requirements for Bibliographic Records, www.frbr.org, 2007.
- [4] I. Buonazia, M. E. Masci, D. Merlitti, *The Project of the Italian Culture Portal and its Development. A Case Study: Designing a Dublin Core Application Profile for Interoperability and Open Distribution of Cultural Contents*, ELPUB2007. Openness in Digital Publishing: Awareness, Discovery and Access - Proceedings of the 11th International Conference on Electronic Publishing held in Vienna, Austria 13-15 June 2007, Edited by: Leslie Chan and Bob Martens. 393–494, 2007.
- [5] The Dublin Core Metadata Initiative, <http://dublincore.org/>.
- [6] Encoded Archival Description, <http://www.loc.gov/ead/>.
- [7] The CIDOC Conceptual Reference Model, <http://cidoc.ics.forth.gr/>.
- [8] IFLA Study Group, Functional Requirements for Bibliographic Records, www.ifla.org/VII/s13/frbr/frbr.pdf, 1998.
- [9] MARC Standards, <http://www.loc.gov/marc/>.
- [10] Nacionalni centar za digitalizaciju, <http://www.ncd.matf.bg.ac.yu>.

- [11] Preporuka formata meta-podataka za objekte digitalizovane nacionalne baštine (pokretna kulturna dobra), 2007.
<http://www.ncd.matf.bg.ac.yu/?page=news&lang=sr&file=predlogStandardaMetadata.htm>
- [12] SPECTRUM, the UK Museum Documentation Standard,
<http://www.mda.org.uk/spectrum.htm>.
- [13] The European Library, <http://www.theeuropeanlibrary.org/>
- [14] Recommendations for coordination of digitization of cultural heritage in South-Eastern Europe, Conclusions of the Regional Meeting on Digitization of Cultural Heritage, Ohrid, Macedonia, 17-20 March 2005, Review of the National Center for Digitization, 2 -7, 2005.
(<http://elib.mi.sanu.ac.yu/files/journals/ncd/7/ncd07002.pdf>)
- [15] Dečija digitalna biblioteka, <http://digital.nbs.bg.ac.yu/decije/>
- [16] XML Schema (W3C), [http://en.wikipedia.org/wiki/XML_Schema_\(W3C\)](http://en.wikipedia.org/wiki/XML_Schema_(W3C))
- [17] Wolfgang Meier. Index-Driven XQuery Processing in the eXist XML Database. Paper presented at XML Prague 2006, <http://exist.sourceforge.net/xmlprague06.html>
- [18] Dušan Tošić, *XML-tehnologije i digitalizacija*, Review of the National Center for Digitization, 1 -12, 2005, (<http://elib.mi.sanu.ac.yu/files/journals/ncd/3/d001download.pdf>)

Dodatak A

Primer

Primerom koji sledi ilustruje se način na koji je opisana knjiga poznatog srpskog komediografa Branislava Nušića „Autobiografija”, prvo u skladu sa preporukom formata meta-podataka, a zatim i u skladu sa definisanom XML shemom. Ova knjiga pripada kolekciji Dečija digitalna biblioteka [15]. Ilustrator knjige je Dragoslav Živković, dok su urednici izdanja Simeon Marinković i Slavica Marković. Knjigu je izdao Kreativni centar. Digitalizaciju knjige uradio je Nikola Pavlović iz Narodne biblioteke Srbije. Autor meta-podataka je Tamara Butigan.

Tabela A.1 sadrži podatke o autoru. Tabele A.2, A.3 i A.4 sadrže podatke o ilustratoru i urednicima. Tabela A.5 sadrži podatke o digitalnoj verziji knjige. Tabela A.6 sadrži podatke kojima se opisuje štampano izdanje knjige.

Descriptive	
Name	
* First_name	Branislav
* Family_name	Nušić
Name_version	
* First_name	Ben
* Family_name	Akiba
Date_of_birth	1864
Date_of_death	1938
Sex	male
Biography	
* Biography_in_Serbian	http://sr.wikipedia.org/wiki/Бранислав_Нушић
* Biography_in_other_languages	http://en.wikipedia.org/wiki/Branislav_Nusic
Administrative	
Record_Creation_Date	30/3/2007
Record_Creator	T. Butigan
Record_Owner	NBS

Tabela A.1: Osoba – autor

Descriptive	
Name	
* First_name	Dobrosav
* Family_name	Živković
Nickname	Bob
Sex	male
Administrative	
Record_Creation_Date	30/3/2007
Record_Creator	T. Butigan
Record_Owner	NBS

Tabela A.2: Osoba – ilustrator

Descriptive	
Name	
* First_name	Simeon
* Family_name	Marinković
Sex	male
Administrative	
Record_Creation_Date	30/3/2007
Record_Creator	T. Butigan
Record_Owner	NBS

Tabela A.3: Osoba – urednik

Descriptive	
Name	
* First_name	Slavica
* Family_name	Marković
Sex	female
Administrative	
Record_Creation_Date	30/3/2007
Record_Creator	T. Butigan
Record_Owner	NBS

Tabela A.4: Osoba – urednik

Descriptive	
Title	
* Title in Serbian	Digitalizovana dečja knjiga Autobiografija, Branislava Nušića
* Title in English	Digitized children's book: Autobiography written by Branislav Nusic
Creator	
* Identifier	ID (Nikola Pavlović)
* Role	programmer
Location of digital document	http://www.digital.nbs.bg.ac.yu/decije/jpg/nusauto/
Administrative	
Archive date	2004
Digital document format	jpg
Size	17.9MB
Digital document MIME format	image
Capture device	Epson GT 15000
Rights	NBS gained rights from the publishing house
Access rights	Unlimited
Digital object owner	NBS
Record creation date	30/3/2007
Record creator	T. Butigan
Record owner	NBS

Tabela A.5: Digitalni dokument

Descriptive	
Title	
* Original name	Аутобиографија
* Name in Serbian	Autobiografija
* Name in English	Autobiography
Creator	
* Identifier	ID (Branislav Nušić)
* Role	writer
Contributor	
* Identifier	ID (Dobrosav Živković)
* Role	illustrator
Contributor	
* Identifier	ID (Simeon Marinković)
* Role	editor
Contributor	
* Identifier	ID (Slavica Marković)
* Role	editor
Classification	
* Classification_scheme	UDK
* Classification identifier	821.163.41-93
Category	
* National category	non categorized
Description	
* Description_in_Serbian	Autobiografija poznatog srpskog pisca Branislava Nušića ispričana u formi šaljivog romana. Sadrži i kraća objašnjenja manje poznatih reči.
* Description_in_English	Autobiography of famous Serbian writer Branislav Nusic but written as a funny story. Vocabulary sections included.
Date_and_provenance	
* Date of origin	2001
* Provenance of origin	Belgrade
* Coordinates	N 44°47'33", E 20°27'53"
* Version of the cultural monument	First issue
Dimension	
* Dimension name	Number of pages

* Dimension value	283
Dimension	
* Dimension name	Book back height
* Dimension value	24 cm
Type	text
Acquisition	
* Type of acquisition	Legal deposit
* Date of acquisition	06/2001
Owner	NBS
Source object ID	II449580
Administrative	
Rights	Publishing house Kreativni centar
Access rights	Unlimited
Record creation date	30/01/2007
Record creator	T. Butigan
Record owner	NBS
Extensions for standard edition	
Publisher	Kreativni centar
Language	scc
Edition	Collection Pustolovine
Identifier	ISBN 86-7781-042-0

Tabela A.6: Digitalizovano kulturno dobro - klasično izdanje