

# MODELIRANJE AUTORIZIRANIH PODATAKA U OPĆEM OKVIRU ZA OPISIVANJE RESURSA<sup>1</sup>

**Predrag Perožić**

OŠ „Fran Krsto Frankopan“,  
Krk, Hrvatska

## KLJUČNE RIJEČI:

*Ontologija autoriziranih podataka (OAP), format UNIMARC za pregledne zapise, modeliranje autoriziranih podataka, RDF, SKOS, SKOSXL*

## SAŽETAK

*U radu se prikazuje istraživanje najprikladnije mogućnosti predstavljanja bibliografskih autoriziranih metapodataka pohranjenih u IFLA-inu UNIMARC formatu za pregledne zapise (UNIMARC/A) kao semantički povezanih i strojno razumljivih podataka pomoću podatkovnog modela koji nazivamo Opći okvir za opisivanje resursa (**Resource Description Framework – RDF**). Rezultat istraživanja konkretan je proizvod pod nazivom Ontologija autoriziranih podataka – OAP, koja se zasniva na apstraktnoj sintaksi podatkovnog modela SKOSXL/RDF i na semantici koja je modelirana prema specifičnim logičkim podatkovnim elementima formata UNIMARC/A koji opisuju bibliografske entitete. Budući da je model OAP jednostavniji alat od formata UNIMARC/A koji priopćava iste informacije, možemo reći da je postignuta svojevrsna optimizacija.*

*U metodološkom smislu OAP kombinira induktivni (**bottom-up**) i deduktivni (**top-down**) pristup i pokušava postići ravnotežu između: a) jasnoće i konzistentnosti u definiranju bibliografskih entiteta kao OAP klasa i njihovih semantičkih odnosa kao OAP svojstava i b) preciznosti i količine informacija koje opisuju te entitete u skupu autoriziranih podataka.*

1 Rad je nastao manjom preradbom poglavlja 1.1 Uvod, 2. Metodologija, plan i cilj istraživanja, 3.1 Osnova ontologije autoriziranih podataka i 6. Zaključak autorove doktorske disertacije *Modeliranje autoriziranih podataka u Općem okviru za opisivanje resursa*, izrađene pod mentorstvom prof. dr. sc. Mirne Willer i obranjene na Sveučilištu u Zadru 15. travnja 2016. (Perožić, 2016).

## Uvod

Općenito se vjeruje da se svijet sastoji od različitih vrsta stvari. To je tako očigledna pretpostavka da se smatra kako je ne treba dokazivati. Otprilike kao i temeljna pretpostavka da svijet uopće postoji. Za znanost to nije upitno. Kad ne bi bilo tako, ne bi se imalo što istraživati. Iako filozofija u svojim radikalnim pristupima propituje najočiglednije, pa tako i ove pretpostavke, ljudi nisu nikada bili ozbiljno zabrinuti oko toga jesu li stvari uopće ili nisu, već kakve su i koliko ih ima. Možemo reći da je to pitanje svih pitanja, u filozofiji, u znanosti i u svakodnevnom životu.

Pod stvarima ovdje ne mislimo samo na fizičke predmete, već na sve entitete o kojima možemo govoriti: slonove, slova, ljude – žive i mrtve, ideje, emocije, glazbene kompozicije – ukratko, bilo što. Uz to što vjerujemo da ima mnoštvo vrsta stvari, vjerujemo da stvari koje naseljavaju naš svijet imaju svojstva. To je druga neupitna činjenica. Kad stvari ne bi imale svojstva, tj. karakteristike ili atribute, ne bismo o njima mogli misliti i govoriti. Stvar bez svojstava je izvan dosega našeg mišljenja i jezika. Daleko nebesko tijelo za koje znamo samo da postoji, zato jer svijetli, ili šumi, već u tim prvim evidencijama skriva mnoštvo svojstava. Treba ih samo prepoznati. Ako samo postojanje i ne smatramo svojstvom činjenica da neki entitet postoji, dovoljno je da mu možemo dodijeliti ime – za ljudsku spoznaju i komunikaciju to je prvo i najbitnije svojstvo!

Poznavanje svojstava stvari i odnosa za koje vjerujemo da vladaju među stvarima, i među njihovim svojstvima, čini ljudsko znanje. Zato se sustavi koji nam pomažu da stvari, svojstva i njihove odnose dovedemo u red nazivaju sustavima za organizaciju znanja (*knowledge organization systems*, KOS). Drugi naziv, koji je uobičajen u informacijskoj znanosti, posebno u knjižničnoj teoriji, jest nadzirani strukturirani rječnici (*controlled structured vocabularies*).

Nadzirani strukturirani rječnici jesu skup raznorodnih leksičkih proizvoda poput tezaurusa, pojmovnika, klasifikacijskih sustava, popisa predmetnih odrednica, taksonomija i njihovih raznih kombinacija koji se međusobno razlikuju po stupnju svoje strukturiranosti i po vrsti nadzora koji se nad njima provodi. U knjižničnim bazama podataka ovi se rječnici tradicionalno nazivaju normativne datoteke ili datoteke preglednih zapisa. Izraz za takve skupove, preveden prema engleskom, jest autorizirani metapodaci (*authority metadata*).

Osnovni zadatak arhivske, knjižnične i muzejske djelatnosti sastoji se u tome da se određene vrste stvari o kojima se brinu ove ustanove, a koje ćemo skraćeno zvati građa, na standardizirani način opišu i pohrane kako bi se poslije mogle

relativno brzo i jednostavno pronaći. Digitalizacija građe i podataka o građi te mogućnost mrežnog pristupa digitalnim sadržajima skraćuje vrijeme pronalazanja i pristupa građi na minute ili čak sekunde. Time se značenje i uloga sustava za organizaciju znanja što postoje već stoljećima nije umanjilo, naprotiv, pred tim je sustavima izazov razvoja i prilagođavanja tehnologijama semantičkog *weba* kako bi se osiguralo da brzina kretanja kroz univerzum informacija dobije željeni rezultat: razumljivu i korisnu informaciju.

Misija semantičkog *weba* mogla bi se opisati kao izgradnja računalnih aplikacija za povezivanje podataka na temelju značenja kako bi se informacije na *webu* pronalazile brzo, precizno i lako. Informacijski stručnjaci iz baštinske domene, s jedne strane, i *web* znanstvenici, s druge strane, prepoznali su zajedničke ciljeve pa možemo reći da najstarija i najmlađa škola za organizaciju informacija sve konstruktivnije surađuje (*Library Linked Data Incubator Group Final Report*, 2011).

Sve dok se bavimo relativno malim brojem stvari, spomenuti zadaci njihova opisivanja i pronalaznja nisu bitan teorijski ni praktični problem. Međutim, kad se suočimo s izazovom uvođenja reda među milijunom raznolikih artefakata, poput tekstova, audiovizualnih proizvoda i realija u fizičkom i digitalnom obliku, posao se komplicira i zahtijeva ne samo složena tehnološka rješenja nego i odgovore na neka temeljna filozofska i logička pitanja, kao što su klasifikacija entiteta, definicija njihovih odnosa i svojstava, mogućnost izvođenja strojnog zaključivanja itd. U radu se predlaže moguće rješenje za obje razine problema koje nazivamo modeliranjem autoriziranih metapodataka u Okviru za opisivanje resursa (*Resource Description Framework – RDF*) (*RDF Primer*, 2004; *RDF 1.1 Primer*, 2014). Preciznije, zanimat će nas Jednostavni sustav za organizaciju znanja (*Simple Knowledge Organization System – SKOS*) (*SKOS Primer*, 2009; *SKOS System Reference*, 2009) koji je jedna od prvih standardiziranih RDF aplikacija. Rezultat je tog modeliranja Ontologija autoriziranih podataka (OAP) koja se zasniva na podacima pohranjenim u IFLA-inu formatu UNIMARC za pregledne zapise (*UNIMARC Authorities Format – UNIMARC/A*) (*UNIMARC Authorities*, 2009). Modeliranje podataka u RDF podatkovnom modelu znači predstavljanje svih podatkovnih elemenata pomoću dviju osnovnih kategorija RDF podatkovnog modela, RDF klasa i RDF svojstava.

## Metodologija, plan i cilj istraživanja

### *Metodologija modeliranja autoriziranih metapodataka*

Ovaj rad bavi se analizom i modeliranjem autoriziranih metapodataka za imena bibliografskih entiteta s aspekta njihova prikazivanja standardima semantičkog *weba* u svrhu prijedloga dizajna ontološkog rječnika koji se smatra najpogodnijim za modeliranje autoriziranih podataka iskazanih formatom UNIMARC za pregledne zapise. U skladu s tim, metodologija istraživanja zasniva se na analizi dokumenata, standarda i formalnih specifikacija bibliografskih modela i formata te standarda semantičkog *weba*.

Dizajn predloženog ontološkog rječnika strukturno je sličan ontološkim rječnicima poput DCMI izrazi (DCMI terms, 2012), odnosno MADS/RDF ontologije (MADS/RDF, 2015). Rječnik Ontologije autoriziranih podataka (OAP) sastoji se od dvaju skupova elemenata, specifikacije OAP klasa i specifikacije OAP svojstava. Opis OAP klasa uključuje vrijednosti sljedećih atributa:

1. formalni naziv ili identifikator (IRI klase)
2. neformalni naziv klase (rdf:label)
3. objašnjenje (rdfs:comment)
4. instancija od (rdf:type)
5. najbliža nadklasu (rdfs:subClassOf)
6. subjekt za svojstvo (popis svojstava kojima je klasa definirana kao subjekt)
7. objekt za svojstvo (popis svojstava kojima je klasa definirana kao objekt)
8. disjunktna klasa (owl:disjointWith)
9. korelacija s UNIMARC/A podatkovnim elementima.

Opis OAP svojstava uključuje vrijednosti sljedećih atributa:

1. formalni naziv ili identifikator (IRI svojstva)
2. neformalni naziv svojstva (rdf:label)
3. objašnjenje (rdfs:comment)
4. instancija od (rdf:type)
5. najbliže nadsvojstvo (rdfs:subPropertyOf)
6. subjekt klasa (rdfs:domain, tj. popis klasa koje su definirane kao subjekt svojstva)
7. objekt klasa (rdfs:range, tj. popis klasa koje su definirane kao objekt svojstva)
8. inverzno svojstvo (owl:inverseOf)

## 9. korelacija s UNIMARC/A podatkovnim elementima.

Praktični dio istraživanja sastoji se u modeliranju podataka koji su pohranjeni u formatu UNIMARC/A kako bi se došlo do konkretnog rječnika koji ima opisani dizajn. Bitno je razlikovati modeliranje formata od modeliranja podataka koji su sadržaj tog formata i koje modeliramo na temelju semantičkih, odnosno logičkih podatkovnih elemenata koje pronalazimo u njihovu opisu. Stvarni semantički, odnosno logički elementi opisa bibliografskih entiteta ne moraju se uvijek podudarati s podatkovnim elementima koji su definirani u formatu UNIMARC/A, odnosno s njegovim označiteljima sadržaja koji su predstavljeni poljima i potpoljima.

Format UNIMARC/A ima svoju strukturu, možemo reći svoju sintaksu, koju u postupku modeliranja možemo mehanički slijediti na način da svaki UNIMARC/A označitelj sadržaja direktno pretvaramo u odgovarajući RDF element. Takva metoda modeliranja naziva se *bottom-up* metodom i u suštini predstavlja modeliranje formata (Willer, Dunsire, Perožić, 2013).<sup>2</sup> Njezin cilj nije izgradnja konzistentne ontologije, već izgradnja alata koji služi za ekstrakciju podataka iz postojećeg formata bez gubitaka informacija. *Bottom-up* metoda ne bavi se analizom semantičkog sadržaja u postojećoj normativnoj bazi podataka i problemom njegova artikuliranja u kontekstu novog podatkovnog modela. Smatra se da je mehanički prijenos podataka istovremeno i prijenos informacija.

Za razliku od modeliranja formata, modeliranje podataka zapravo je artikuliranje semantičkog sadržaja pohranjenog u formatu. Središte interesa jest skup autoriziranih metapodataka koji promatramo kao sustav za organizaciju znanja. Ovaj pristup podrazumijeva analizu spomenutih semantičkih, odnosno logičkih elemenata opisa koje nalazimo u preglednim zapisima i njihovo eksplicitno i konzistentno definiranje pomoću RDF elemenata. Takva metoda modeliranja naziva se *top-down* metodom.

OAP je svojevrsna kombinacija ovih dviju metoda, ali ponajviše se rukovodi *top-down* pristupom. Primjerice, u ovom se radu smatra da UNIMARC/A polja iz bloka odrednica 2--, odnosno kombinacija tih polja i indikatora, može poslužiti za identifikaciju entiteta koji se opisuje u određenom preglednom za-

2 Vidjeti modeliranje RDF skupova elemenata formata UNIMARC za bibliografske podatke u Otvorenom registru metapodataka (Open Metadata Registry: OMR) na: <http://metadataregistry.org/schema/list/page/9.html>. Isti pristup u modeliranju RDF skupova elemenata za MARC21 format, gdje su deklarirani svi označitelji sadržaja u formatu razvrstani u liste prema blokovima podataka (00X, 0XX, 1XX, 2XX, itd.). Svaki označitelj sadržaja mehanički je predstavljen jednim RDF svojstvom bez oblikovanja logičkog okvira ili semantike, vidjeti na: <http://metadataregistry.org/schema/list/page/3.html> (23-01-2019).

pisu, ali se to načelo ne slijedi mehanički po pravilu – jedno polje jedan entitet. Kako bi se izgradila konzistentna ontologija za opisivanje autoriziranih metapodataka, deklariraju se dodatni entiteti u obliku OAP klasa, kao što je `oap:ObjektStvarnogSvijeta`, koji nije formalno predstavljen strojno razumljivim UNIMARC/A označiteljima, ili kao što je `oap:ZamijenjeniAutoriziraniEntitet`, koji u formatu nije dovoljno jasno i eksplicitno strukturiran pa se taj nedostatak nadoknađuje modeliranjem OAP elementima.

Treba uočiti da UNIMARC/A označitelji sadržaja, koji se sastoje od polja xxx + indikatora xx + potpolja \$\_, u različitim blokovima podataka često ponavljaju isti podatkovni element. Primjerice, „xxx-xx-\$a Prvi element unosa“ označitelj je koji čuva mjesto za prvi dio pristupnice, tj. strukturirane leksičke oznake. Kao takav ponavlja se u većini polja bloka 2--, bloka 4--, bloka 5-- i bloka 7-- . Što je prvi dio pristupnice, format ne definira, već to prepušta standardima za opisivanje imena bibliografskih entiteta, u pravilu nacionalnim kataložnim pravilima.

Čuvari mjesta (*place holders*) za određenu vrijednost jesu varijable što su u osnovi sintaktički, a ne semantički elementi. Format UNIMARC/A dizajniran je kao sintaktička struktura koja sadrži velik broj varijabli kako bi se mogao prilagoditi različitim lokalnim upotrebama koje će definirati vrijednost varijabli i pretvoriti ih u konstante (definirane vrijednosti). OAP se zasniva na uvjerenju da je najveći broj UNIMARC/A varijabli, posebice u slučaju podatkovnih elemenata koji opisuju pristupnice, moguće definirati kao konstante i na taj način sintaktičke elemente pretvoriti u semantičke elemente. Dakle, kao što se entiteti koji su predmet opisa preglednih zapisa modeliraju kao OAP klase (`oap:OsobnoIme`, `oap:ZaštitniZnak`, `oap:PolitičkaZemljopisnaJedinica` itd.) i dijelovi njihovih leksičkih oznaka koje su njihova imena i nazivi, također se modeliraju kao OAP klase koje čine točno određenu vrstu podataka (`oap:ElementOsobnoIme`, `oap:ElementObiteljskoIme`, `oap:ElementGrad`, `oap:ElementDržava`, `oap:ElementZaštitiniZnak` itd.). Na taj način varijable poput „prvi element unosa“, „preostali dio naziva“ i slično, pretvaramo u klase definiranih podataka.

Mogli bismo reći da se osnovno oblikovno načelo OAP ontologije sastoji u tome da se:

- a) UNIMARC/A označitelji sadržaja koji opisuju dijelove pristupnice (blokovi podataka 2--, 4--, 5-- i 7--) i podaci o entitetima koji su strukturirani kao pristupnice (polja 640 i 341) modeliraju kao OAP klase čime se semantizira OAP rječnik, integriraju podatkovni elementi i smanjuje njihov ukupni broj

- b) UNIMARC/A označitelji sadržaja koji ne opisuju pristupnice, tj. strukturirane leksičke oznake, nego karakteristike entiteta koje se odnose na objekt stvarnog svijeta, upotrebu entiteta, klasifikaciju, izvor informacija o entitetu itd. (blokovi podataka 1--, 3--, 6--, 8--) u pravilu modeliraju kao OAP svojstva.

Na koncu, skupina podataka koji ne opisuju bibliografski entitet, nego UNIMARC zapis kao takav dodatna je meta razina opisa, tj. podaci o autoriziranim metapodacima. To su podaci koji se nalaze u bloku podataka 0--, 1-- i 8--, a ovdje ih nazivamo administrativnim metapodacima (ADM). RDF elementi za opisivanje zapisa deklarirani su kao zaseban rječnik u posebnom imenskom prostoru.

### ***Plan istraživanja ili postupci u modeliranju autoriziranih metapodataka***

Modeliranje autoriziranih metapodataka u ovom radu sastoji se od sljedećih postupaka:

1. Oblikovanje hipotetičkog imenskog prostora za skup autoriziranih podataka koji su pohranjeni u format UNIMARC/A. Kako bi se mogli konstruirati funkcionalni primjeri preglednih zapisa u RDF/XML notaciji koji se mogu testirati u RDF validatoru (W3C RDF Validation Service), potreban je sintaktički valjani IRI (Internationalized Resource Identifier) za skup autoriziranih podataka. Primjeri koji se izrađuju na temelju normativne baze podataka informacijskog sustava CROLIST imaju osnovni IRI **<http://www.crolist/autpod/>**. IRI pojedinačnog entiteta konstruira se tako da se na osnovni dio dodaje prefiks **i** za imenske, tj. autorske pristupnice, odnosno prefiks **t** za tematske, tj. predmetne pristupnice. Iza prefiksa navodi se devet znamenki koje se generiraju iz identifikatora zapisa u polju 001. Primjerice **<http://www.crolist/autpod/i123456789>**. Ako se prikazuje primjer iz neke druge normativne datoteke rabit će se oblik **<http://www.xxx/autpod/>**.
2. Oblikovanje hipotetičkog imenskog prostora za deklariranje elemenata OAP ontologije. Kako bi se mogli konstruirati funkcionalni primjeri preglednih zapisa u RDF/XML notaciji koji se mogu testirati u RDF validatoru, potreban je sintaktički valjani IRI za identifikaciju elemenata (klasa i svojstva) OAP rječnika. Također, ovaj imenski prostor potreban je za sastavljanje strojno čitljivih aksioma OAP ontologije u OWL/XML notaciji. IRI glasi **<http://www>**.

**unimarca/oap/v1#.**

3. Oblikovanje hipotetičkog imenskog prostora za deklariranje elemenata ADM ontologije, tj. rječnika za predstavljanje administrativnih podataka ili podataka koji opisuju UNIMARC/A zapis. Kako bi se mogli konstruirati funkcionalni primjeri cjelovitih preglednih zapisa u RDF/XML notaciji koji se mogu testirati u RDF validatoru, potreban je sintaktički valjani IRI za identifikaciju elemenata (1 klasa + 16 svojstava) ADM rječnika. Također, ovaj imenski prostor potreban je za sastavljanje strojno čitljivih aksioma ADM ontologije u OWL/XML notaciji. IRI glasi **http://www.unimarca/adm/v1#**. ADM skup elemenata importiran je u OAP skup elemenata:
 

```
<owl:Ontology rdf:about="http://www.unimarca/oap/v1#">
<owl:imports rdf:resource="http://www.unimarc/adm/v1#" />
</owl:Ontology>
```
4. Oblikovanje hipotetičkog imenskog prostora za deklariranje koncepata UNIMARC rječnika vrijednosti koji predstavljaju kodirane vrijednosti u bloku kodiranih podataka 1--. Osnovni IRI glasi: **http://www.unimarc/rječnik/**. Treba uočiti da se neki rječnici vrijednosti deklariraju kao dodaci formatu UNIMARC/B (kodirane vrijednosti za zemlje, jezike itd.), ali se upotrebljavaju za opisivanje entiteta i u formatu UNIMARC/A. S druge strane, neki rječnici kodiranih vrijednosti upotrebljavaju se samo za opisivanje specifičnog entiteta u jednom formatu i deklarirani su u sklopu određenog bloka podataka. Ovaj hipotetički imenski prostor konstruira se iz istog razloga kao i prethodno spomenuti imenski prostori, ali rječnici vrijednosti nisu dio skupa elemenata OAP ontologije pa se ne opisuju u OAP specifikaciji. Naime, koncepti rječnika vrijednosti nisu OAP klase ni OAP svojstva, ali su instancije klase oap:AutoriziraniEntitet i istovremeno instancije klase skos:Concept.
5. Definiranje entiteta koji se opisuju u formatu UNIMARC/A i dijelova leksičkih oznaka koje predstavljaju imena i nazive entiteta i njihovo modeliranje kao OAP klase. Osnovni su korpus podaci u bloku 2--, 4--, 5--, 7--, polju 640 i 341.
6. Definiranje atributa i odnosa među entitetima i njihovo predstavljanje OAP svojstvima. Korpus su svi podatkovni elementi u formatu UNIMARC/A.
7. Definiranje dodatnih elemenata (klasa i svojstava) za izražavanje informacija o entitetima koje se nalaze na različitim mjestima u formatu. Korpus su blokovi podataka 1--, 3--, 6-- i 8--.

Analitičkom i kvalitativnom metodom pokazat će se da modeliranje autoriziranih

ranih podataka u OAP rječniku povećava njihov stupanj izražajnosti i njihovu funkcionalnost u kontekstu *weba* podataka. Analitička metoda sastoji se u argumentaciji koja je provedena u glavnom dijelu autorove disertacije (Perožić, 2016) gdje se dijagramima i primjerima RDF/XML dokumenata dokazuje semantička ekvivalentnost OAP/RDF grafova s UNIMARC/A zapisima. Konačni kvalitativni test jest validacija cjelovitih UNIMARC preglednih zapisa u RDF/XML notaciji u RDF validatoru.

### ***Cilj istraživanja***

Primarni je cilj istraživanja izgradnja OAP rječnika za predstavljanje autoriziranih metapodataka u bibliografskoj domeni čime se bibliografski sustavi za organizaciju znanja stavljaju u kontekst *weba* podataka. Podsjetimo se da se OAP zasniva na sintaksi SKOSXL/RDF podatkovnog modela i na semantici koja je modelirana prema stvarnim logičkim podatkovnim elementima formata UNIMARC/A.

Sekundarni je cilj istraživanja primjena programa univerzalnog bibliografskog nadzora (Universal Bibliographic Control – UBC) na autorizirane metapodatke u kontekstu semantičkog *weba*. To znači da se UBC program realizira u okviru općeg načela semantičkog *weba* koje glasi: „bilo tko može reći bilo što o bilo čemu“ (*Anybody can say Anything about Any thing*, AAA). Načelo 3A podrazumijeva slobodu kreiranja metapodataka i lokalnih rječnika pod uvjetom da su rječnici:

- a) modelirani prema zajedničkoj apstraktnoj sintaksi ili podatkovnom modelu RDF
- b) iskazani u jednoj od standardnih strojno čitljivih konkretnih sintaksi (RDF/XML, Turtle, JSON, N-Triples itd.)
- c) mapirani, tj. povezani eksplicitnim semantičkim odnosima s općeprihvaćenim standardnim globalnim rječnicima i jezicima za predstavljanje znanja (RDF, RDFS, SKOS, OWL).

Ispunjavanje a) i b) uvjeta omogućuje interoperabilnost rječnika, dok ispunjavanje uvjeta pod c) omogućuje određenu razinu automatskog zaključivanja.

Možemo reći da je u ovom novom kontekstu stara paradigma UBC-a koja glasi *jedan entitet – jedna pristupnica* zamijenjena načelom *jedan (konceptualni) entitet – mnoštvo semantički povezanih leksičkih oznaka*.

Rezultat istraživanja u ovom radu jest ontologija koja s relativno malim skupom elemenata (102 klase i 147 svojstava) izražava sve informacije koje nalazimo u postojećem formatu UNIMARC/A te osigurava elemente za izražavanje dodatnih semantičkih odnosa u kontekstu semantičkog *weba* koji u formatu UNIMARC/A nisu predviđeni. Budući da je OAP rječnik neovisan o podatkovnom formatu, može se serijalizirati u različitim RDF notacijama, ovisno o tehnološkim mogućnostima i potrebama korisnika.

Konačna namjena sustava za organizaciju znanja, odnosno skupova autoriziranih metapodataka u knjižničnim i srodnim informacijskim sustavima, jest da definiraju koncepte i dodjeljuju im standardizirane identifikatore i oznake kako bi njima označivali građu koja govori o određenim stvarima u svijetu, ili koja sama kao takva predstavlja stvari od interesa. Sustavi za organizaciju znanja razvijaju se usporedno s ljudskim znanjem i registriraju taj razvoj na dva načina:

- a) uvođenjem novih koncepata što se odražava kao uvođenje novih leksičkih oznaka i novih httpIRI-ja koji identificiraju koncepte u SKOSificiranim<sup>3</sup> sustavima za organizaciju znanja
- b) zamjenom postojećih koncepata koji se smatraju neadekvatnima što se odražava kao promjena u statusu njihovih leksičkih oznaka i/ili httpIRI-ja.

Kada se određeni koncept iz bilo kojeg razloga više ne upotrebljava kao aktualni koncept koji služi za označivanje građe u OAP ontologiji, smatra se zamijenjenim konceptom i formalno se deklarira kao instancija klase `oap:ZamijenjeniAutoriziraniEntitet` jer njegovu ulogu ili funkciju preuzima jedan ili više drugih koncepata. Važno je razumjeti da se u kontekstu semantičkog *weba* zastarjeli ili bivši koncepti ne bi smjeli brisati iz memorije sustava za organizaciju znanja jer se na taj način stvaraju prekidi u prethodno uspostavljenim semantičkim vezama. Zastigurno postoji više načina da ovakvi entiteti budu obilježeni i arhivirani na strojno razumljivi način. U OAP ontologiji klasa bivših, tj. zamijenjenih koncepata ne smatra se više potklasom od `skos:Concept`, nego potklasom od `owl:Thing`, najopćenitije klase u OWL rječniku koja predstavlja klasu svih individua. Instancija klase `owl:Thing` može biti bilo što. O razlozima za takav dizajn detaljnije se govori u idućem poglavlju.

Da bi autorizirani podaci u okruženju semantičkog *weba* ispunjavali svoju ulo-

3 Sustavi modelirani u skladu sa SKOS-XL/RDF podatkovnim modelom. Nap. ur.

gu, ključna je eksplicitna definicija i identifikacija entiteta koji se opisuju u skupu autoriziranih podataka, a koja se zasniva na semantičkim razinama koje smo u prethodnom odlomku podcrtali. Na temelju toga izvedena je **osnovna hipoteza istraživanja**:

Kako bi se postigla eksplicitna definicija i nedvosmislena identifikacija svih entiteta u skupu autoriziranih podataka, OAP ontologija treba na strojno razumljiv način modelirati četiri semantičke razine – konceptualnu, leksičku, realnu i globalnu.

To će se postići na sljedeći način:

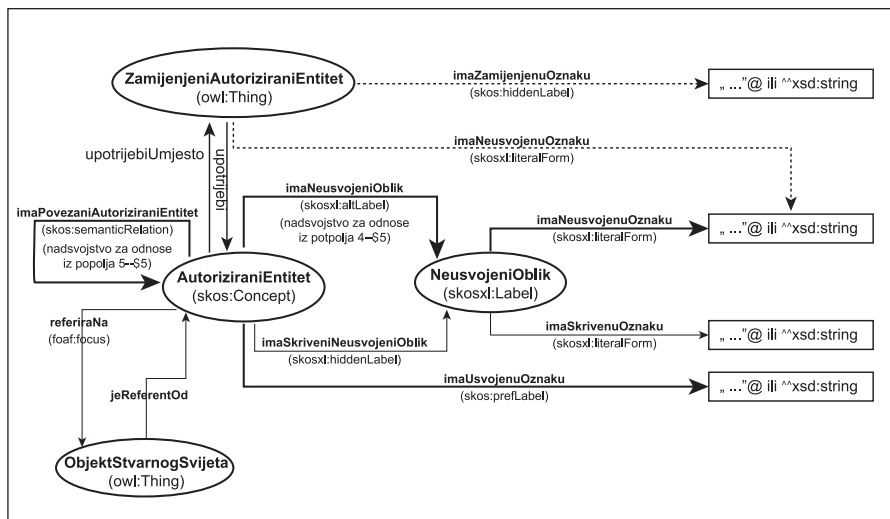
1. OAP klasama eksplicitno će se predstaviti koncepti, leksičke oznake, objekti stvarnog svijeta i zamijenjeni koncepti
2. OAP svojstvima eksplicitno će se izraziti sljedeće vrste odnosa:
  - odnos koncept – koncept
  - odnos koncept – zamijenjeni koncept
  - odnos koncept – oznaka
  - odnos zamijenjeni koncept – oznaka
  - odnos koncept – objekt stvarnog svijeta
  - odnos objekt stvarnog svijeta – koncept.

U idućem poglavlju objasnit će se osnova mehanizama i namjene OAP ontologije čime će se ostvariti cilj istraživanja, a u zaključnom poglavlju odgovorit će se na sljedeća **istraživačka pitanja**:

1. Koji je stupanj izražajnosti Ontologije autoriziranih podataka u odnosu na izražajnost podataka pohranjenih u formatu UNIMARC za pregledne zapise?
2. Koji su konstrukti i mehanizmi iz RDF/SKOS rječnika potrebni kako bi se izrazila sintaktička struktura svih vrsta pristupnica u preglednim zapisima?

## Ontologija autoriziranih podataka – OAP: osnova

Dijagram na Slici 1. prikazuje osnovne elemente Ontologije autoriziranih podataka.



SLIKA 1. Osnova Ontologije autoriziranih podataka

Osnovu ontologije čine **četiri** klase koje su na dijagramu prikazane ovalima: oap:AutoriziraniEntitet, oap:ZamijenjeniAutoriziraniEntitet, oap:NeusvojeniOblik i oap:ObjektStvarnogSvijeta i **jedanaest** svojstava koja su prikazana strelicama. Ispod naziva klase ili svojstva u okrugloj je zagradi navedena najbliža natklasa prema kojoj vrijedi odnos rdfs:subClassOf, odnosno najbliže nadsvojstvo prema kojem vrijedi odnos rdfs:subPropertyOf. Odnosi koji se najviše upotrebljavaju jer opisuju najveći broj podataka u skupu autoriziranih podataka označeni su podebljanim strelicama.

Kvadrati na desnoj strani dijagrama nadzirane su leksičke oznake. To su nizovi znakova iz prirodnog jezika kojima se pridružuje oznaka lingvističkog sustava kojem pripadaju (na dijagramu označeno „...“@xx), a ako lingvistički sustav nije moguće odrediti, leksička oznaka smatra se jednostavno alfanumeričkim nizom znakova (na dijagramu označeno xsd:string). Podsjetimo se da jezičnu oznaku nije moguće strojno generirati kada u UNIMARC/A zapisu nedostaju podaci o jeziku i pismu osnovne pristupnice ili zato što je ukupna pristupnica sastavljena od leksičkih oznaka koje pripadaju različitim jezičnim sustavima. U svakom slučaju, to su vrijednosti koje se izražavaju tipom podataka koji se u RDF

podatkovnom modelu nazivaju literalima koji su formalno predstavljeni klasom `rdf:Literal`. U kontekstu konceptualnog modela FRAD nadzirane leksičke oznake nazivaju se nadziranim pristupnicama (*controlled access points*),<sup>4</sup> a u kontekstu knjižničnog referentnog modela FRBR-LRM (Riva, Žumer, 2015)<sup>5</sup> ove leksičke oznake smatraju se *nomenima (nomen)* što je novodefinirana klasa za sve vrste znakova i/ili simbola po kojima je neki entitet poznat.

Treba napomenuti da struktura i konzistentnost OAP ontologije kao podatkovnog modela ne ovisi neposredno o apstraktnim rješenjima i definicijama koje se donose u konceptualnim bibliografskim modelima. Konceptualni modeli vrlo su općeniti teorijski okvir koji nam pomaže da shvatimo od čega se sastoji domena koju modeliramo. To je pogled s visine. Kategorije konceptualnog modela uglavnom nisu dovoljno precizno i jasno definirane da bi se mogle neposredno upotrebljavati za modeliranje konkretnih podataka. Elementi OAP ontologije nisu formalno pridruženi kategorijama ni jednog konceptualnog modela jer je među njima teško utvrditi striktnu logičke odnose, iako se može govoriti o analogijama i sličnostima. S druge strane, struktura i konzistentnost OAP ontologije neposredno ovisi o definicijama i rješenjima u općim podatkovnim modelima — RDF/S, SKOS/XL te jeziku za izražavanje ontologija na *webu* — OWL (*Web Ontology Language*) (*OWL Primer*, 2012), koji su kreirani s praktičnim ciljem da služe za predstavljanje podataka na *webu* na strojno razumljiv način. To su sintaktičke platforme. Njihovi su elementi precizno definirani i svaka RDF aplikacija može svoje elemente (klase i svojstva) pridružiti na jasan i precizan način (kao potklase i podsvojstva) njihovim elementima (također klasama i svojstvima) čime se jamči minimalna razina interoperabilnosti. Konceptualni bibliografski modeli nam, ovakvi kakvi jesu, to ne jamče.

### ***Osnovni elementi Ontologije autoriziranih podataka***

**Oap: AutoriziraniEntitet** jest klasa konceptualnih resursa čija je leksička oznaka:

- a) usvojena pristupnica opisana u bloku podataka 2-- u UNIMARC/A preglednom zapisu s oznakom stanja zapisa n=novi ili c=ispravljeni (ako se

4 Entiteti i odnosi konceptualnog modela FRAD deklarirani su kao RDF klase i svojstva u Otvorenom registru metapodataka (OMR); vidi model FRAD [http://metadatareregistry.org/schemaprop/list/schema\\_id/24.html](http://metadatareregistry.org/schemaprop/list/schema_id/24.html) (23-01-2019).

5 U trenutku pisanja doktorskog rada FRBR-LRM još nije bio službeno objavljen i nije imao registrirani imenski prostor u kojem bi se formalno deklarirali njegovi elementi (klase i svojstva) (Riva, Žumer, 2015). Model je objavljen 2017. pod naslovom *IFLA Library Reference Model* (Riva, Le Bœuf i Žumer, 2017), ali njegov imenski prostor nije registriran do siječnja 2019.

ispravak ne odnosi na podatke u bloku podataka 2--) ili

- b) bilo koji koncept čija je leksička oznaka jedna od kodiranih oznaka u bloku kodiranih podataka 1--.

Koncepti iz ove druge vrste deklariraju se u zasebnim UNIMARC rječnicima vrijednosti. Treba napomenuti da se rječnici kodiranih vrijednosti pojavljuju u sklopu formata UNIMARC/A i UNIMARC/B, ali u obama slučajevima čine skupove koncepata koji su instancije klase `oap:AutoriziraniEntitet`. Na dnu dijagrama na Slici 1. vidimo da OAP ontologija prikazuje usvojenu pristupnicu kao vrijednost svojstva `oap:imaUsvojenuOznaku` koje je podsvojstvo od `skos:prefLabel`. Na lijevoj strani prikazani su semantički odnosi među autoriziranim entitetima koji se u preglednim zapisima opisuju u bloku podataka 5--. Svojstvo `oap:imaPovezaniAutoriziraniEntitet` podsvojstvo je od `skos:semanticRelation` jer okuplja sve vrste odnosa među konceptima (tradicionalno, srodnih odrednica). Primjećuje se samo u slučaju kada u potpolju 5--\$5 nije naveden specifičan odnos. Svi specifični odnosi kodirani u potpolju 5--\$5 definirani su kao podsvojstva od `oap:imaPovezaniAutoriziraniEntitet`.

**Oap:ZamijenjeniAutoriziraniEntitet** jest klasa bivših, odnosno zamijenjenih autoriziranih entiteta koji su opisani u preglednim zapisima s oznakom stanja zapisa `d=izbrisani` zapis ili `c=ispravljeni` zapis ako se ispravak odnosi na podatke iz bloka 2--. Oznaka „d“ za izbrisani zapis, kao i svaki ispravljeni zapis s oznakom „c“ koji sadrži polje 836, svrstava entitet opisan u takvim zapisima u instanciju klase `oap:ZamijenjeniAutoriziraniEntitet`. Na vrhu dijagrama vidimo da je leksička oznaka zamijenjenog entiteta vrijednost svojstva `oap:imaZamijenjenuOznaku` koja je podsvojstvo od `skos:hiddenLabel`, što znači da se takva oznaka ne prikazuje, ali je aktivna u postupku pretraživanja. Na isti način funkcionira i skrivena oznaka neusvojenog oblika koja uključuje gramatički nepravilne oblike. Valja uočiti da zamijenjeni autorizirani entitet može, ali i ne mora, biti premješten u neusvojene oblike (tradicionalno, opće uputnice). To je odluka koja uglavnom ovisi o kataložnim pravilima i praksi ustanove, a obje su opcije prikazane isprekidanim strelicama čime se htjelo istaknuti da su ovi odnosi međusobno isključivi: 1. opcija — ako je zamijenjeni koncept u formatu UNIMARC/A izbrisan, onda je u OAP modelu njegova leksička oznaka skrivena (vrijednost svojstva `oap:imaZamijenjenuOznaku`), 2. opcija — ako je leksička oznaka zamijenjenog koncepta u formatu UNIMARC/A premještena u uputnice u bloku 4--, onda taj zamijenjeni koncept u OAP modelu postaje instancija klase `oap:NeusvojeniOblik`, a njegova je leksička oznaka vrijednost svojstva

`oap:imaNeusvojenuOznaku`. Naime, instance klase `oap:NeusvojeniOblik` također su instance klase `skosxl:Label` koja je disjunktna s klasom `skos:Concept`. Stoga i klasa `oap:ZamijenjeniAutoriziraniEntitet` mora biti disjunktna s klasom `skos:Concept`. Drugim riječima, bivši koncepti više ne mogu biti koncepti. Zato je najjednostavnije rješenje da se klasa `oap:ZamijenjeniAutoriziraniEntitet` definira kao potklasa najopćenitije klase `owl:Thing` koja uključuje sve vrste entiteta. Možemo zaključiti da su `oap:AutoriziraniEntitet` i `oap:ZamijenjeniAutoriziraniEntitet` međusobno disjunktne klase, ali `oap:ZamijenjeniAutoriziraniEntitet` i `oap:NeusvojeniOblik` nisu disjunktne klase (moguće su individue koje pripadaju jednoj i drugoj klasi). Na koncu ostaje za reći da se odnos aktualnog i zamijenjenog entiteta opisuje inverznim svojstvima `oap:upotrijebi` i `oap:upotrijebiUmjesto`.

**Oap:NeusvojeniOblik** je klasa leksičkih resursa prvog reda i zato je definirana kao potklasa od `skosxl:Label`. Neusvojeni oblici alternativni su oblici autoriziranih entiteta koji su opisani u preglednim zapisima (vrsta zapisa `x`) u bloku podataka 4-- ili u zapisima općih uputnica (vrsta zapisa `y`) u bloku podataka 2--. Modeliramo ih kao resurse prvog reda jer se pojavljuju kao subjekti RDF iskaza. Svojstvo `oap:imaNeusvojeniOblik` poveznica je između autoriziranog entiteta, odnosno koncepta i njegovih alternativnih, tj. neusvojenih oblika. To je nadsvojstvo za sve odnose kodirane u potpolju 4--\$5 i primjenjuje se samo u slučaju kada u potpolju 4--\$5 nije naveden specifičan odnos. `Oap:NeusvojeniOblik` povezan je s leksičkom oznakom svojstvom `oap:imaNeusvojenuOznaku` koje je podsvojstvo od `skosxl:literalForm`. Skriveni neusvojeni oblici i njihove skrivene oznake prikazane su na dijagramu ispod toga i služe za opisivanje i nadzor gramatički nepravilnih oblika koji se ne prikazuju, ali su korisni u postupcima pronalaženja podataka.

**Oap:ObjektStvarnogSvijeta** je klasa entiteta iz stvarnog svijeta na koje referira autorizirani entitet kao njihova konceptualizacija, tj. njihova apstrakcija ili opis. To je odnos koji se izražava svojstvom `oap:referiraNa`. Inverzno svojstvo `oap:jeReferentOd` predviđeno je za povezivanje objekta s njegovim konceptualizacijama u sustavu autoriziranih podataka. Naime, česta je situacija da postoji više koncepata (autoriziranih entiteta) koji referiraju na isti stvarni objekt. Primjerice, autorizirani pseudonim i pravo ime, raniji i kasniji autorizirani naziv itd. Različiti koncepti koji referiraju na isti objekt mogu pripadati različitim sustavima za organizaciju znanja (instancijama klase `oap:SustavKoncepta`). Primjer su toga koncepti iz predmetnog sustava i koncepti iz imenskog (autorskog) skupa podataka koji se mogu razlikovati po obliku i opisu, ali referiraju na isti objekt. U praksi klasa `oap:ObjektStvarnogSvijeta` najčešće uključuje stvarnu osobu (živu ili umrlu), organizaciju ili grupu ljudi, ma-

terijalni predmet te mjesto ili događaj u realnom prostoru i vremenu. Budući da autorizirani podaci nisu namijenjeni opisivanju objekata stvarnog svijeta, OAP ontologija ne bavi se detaljnom klasifikacijom takvih objekata. Autorizirani podaci opisuju koncepte i njihove leksičke oznake, a objekti stvarnog svijeta rabe se samo kao pomoć u identifikaciji i definiranju koncepata. Stoga se očekuje da instance klase `oap:ObjektStvarnogSvijeta` budu preciznije opisane pomoću specifičnih klasa iz drugih rječnika, primjerice `foaf:Person`, `foaf:Group`, `foaf:Organization` (FOAF, 2014), `dct:Event`, `dct:Location`, `dct:PhysicalObject`, `dct:PeriodOfTime` itd. Treba istaknuti da se entiteti FRBR modela *djelo* i *izraz* koji se u trećem izdanju formata UNIMARC/A eksplicitno mogu upotrebljavati u opisivanju naslova u polju 230, također trebaju smatrati vrstom objekta stvarnog svijeta. S obzirom na to da su entiteti *djelo* i *izraz* u formatu UNIMARC/A kodirane vrijednosti u polju 154, modelirat će se kao koncepti u odgovarajućem UNIMARC/A rječniku vrijednosti i upotrebljavati kao atributi odgovarajuće instance klase `oap:ObjektStvarnogSvijeta`.<sup>6</sup> Osnovna je namjena elementa `oap:ObjektStvarnogSvijeta` da se autorizirani podaci koji se odnose na stvarni svijet (mjesto i datum rođenja i smrti autora, adresa korporativnog tijela, mjesto prvog izvođenja djela itd.) na strojno razumljiv način razdvoje od podataka o konceptima i leksičkim oznakama. Podsjetimo se da je eksplicitna identifikacija entiteta uvjet izvođenja konzistentnog automatskog zaključivanja. Na koncu, na taj se način olakšava nedvosmisleno povezivanje s entitetima iz drugih RDF skupova podataka koji opisuju artefakte iz stvarnog svijeta u oblaku povezanih podataka (*Linked data cloud*).

Treba uočiti da podaci navedeni u zapisima općih objasnidbenih uputnica (vrsta zapisa *z*) nisu modelirani u OAP ontologiji. Razlog je tomu što izrazi koji se navode u tim zapisima nisu cjelovita imena ili nazivi određenog entiteta, najčešće su to samo prefiksi imena („De la...“) ili početak fraze koji čini tipičan naslov („Konferencija...“) i slično. Objasnidbeni zapisi savjeti su ljudskim korisnicima za oblikovanje upita i pronalaženje određenih oblika pristupnica ili pak objašnjavaju određena kataložna pravila koja se primjenjuju u skupu autoriziranih podataka.

6 U rječniku vrijednosti u kojem se deklariraju, ovi se koncepti mogu povezati odnosom `oap:vidi_i` (podsvojstvo od `rdf:seeAlso`) s klasama *work* i *expression* iz konceptualnog modela FRBR koji je formalno deklariran u Otvorenom registru metapodataka (OMR) [http://metadataregistry.org/schemaprop/list/schema\\_id/5.html](http://metadataregistry.org/schemaprop/list/schema_id/5.html) (23-01-2019). Budući da su elementi *djelo* i *izraz* u UNIMARC/A rječniku modelirani kao SKOS koncepti, a *work* i *expression* u OMR-u kao OWL klase, formalno gledano, između njih nije valjano tvrditi odnos `owl:equivalentClass` (semantička ekvivalenost OWL klase), ni `skos:exactMatch` (semantička ekvivalenost SKOS koncepata) ni `owl:sameAs` (identitet individua), iako su i jedni i drugi formalizacija istih entiteta iz konceptualnog modela FRBR.

## Zaključak

Ontologija autoriziranih podataka (OAP) inspirirana je MADS/RDF ontologijom (MADS/RDF) koja u modeliranju autoriziranih podataka odabire deduktivni ili tzv. *top-down* pristup. To je pristup u kojem se domena ili dio svijeta na koji se ontologija odnosi pokušava logički konzistentno i sveobuhvatno opisati pomoću RDF klasa i svojstava koje nazivamo skupom elemenata ontologije. Kako je u uvodnom dijelu već spomenuto, mnoštvo vrsta stvari temeljno je obilježje svijeta u cjelini, pa tako i svakog njegova dijela. Stoga je jedna od osnovnih karakteristika ontologija s deduktivnim pristupom relativno velik broj klasa koje predstavljaju entitete, odnosno različite vrste individua za koje se smatra da postoje ili „naseljavaju“ domenu koja se opisuje. Definiranje klasa i njihovih logičkih odnosa zahtjevan je posao koji većina konstruktora računalnih aplikacija za semantičku obradu podataka pokušava izbjeći. Međutim, ako je taj posao uspješno izveden, kao rezultat imamo logički uređen sustav koji omogućuje izvođenje automatskog zaključivanja i generiranje dodatnog, eksplicitnog znanja koje ljudi vjerojatno ne bi nikada deducirali zbog nemogućnosti promišljanja takve količine podataka. U načelu, što je veći broj klasa koji je deklariran u nekoj ontologiji, manji je broj svojstava koji je potreban za opisivanje domene jer se izražajnost ovih dvaju tipova elemenata međusobno nadoknađuje. Naime, ako klase relativno precizno opisuju i iscrpno dijele domenu, onda svojstva ne moraju biti naročito specifična i brojna, i obrnuto.

Za razliku od toga, induktivni ili tzv. *bottom-up* pristup izbjegava eksplicitno logičko strukturiranje domene i ostavlja ga više-manje otvorenim pitanjem, ili pak implicitno prihvaća neki od postojećih konceptualnih modela koji na vrlo općenit način strukturiraju domenu na koju ontologija referira. Primarna je karakteristika takve ontologije minimalan broj klasa, u krajnjem slučaju to može biti samo jedan nediferencirani entitet te relativno velik broj svojstava kojima se nadoknađuje potrebna izražajnost skupa elemenata. Broj svojstava izrazito induktivne ontologije može dostići brojku od više tisuća i takva su svojstva vrlo specifična. Primjer je takvog pristupa modeliranje skupa elemenata za opisivanje bibliografskih podataka koji su pohranjeni u formatu UNIMARC/B, projekt čiji su autori Gordon Dunsire i Mirna Willer, a suradnik na projektu autor je ovog rada (Dunsire, Willer, Perožić, 2013).<sup>7</sup> Skup elemenata u spomenutom projektu ima jedanaest tisuća svojstava koja su konstruirana na način da se iz svakog UNIMARC/B označitelja sadržava

7 Vidi i raspravu o metodološkom pristupu u ovom projektu (Perožić, 2012) te općenito o bibliografskim podacima na *webu* podataka (Willer i Dunsire, 2013).

ja jednostavnom transformacijom konstruira jedno RDF svojstvo. Kao subjekt tih svojstava pojavljuje se samo jedna klasa koja nema preciznu definiciju, uglavnom iz prije spomenutog razloga, da se izbjegne klizav teren logičkog modeliranja i klasifikacije bibliografskih resursa koje ta klasa uključuje. Mogli bismo reći da je to svojevrсна metaklasa koja okuplja sve vrste građe koja se opisuje u bibliografskim zapisima. Spomenut ćemo samo da je jedna od mogućih interpretacija da se na poziciju te klase postavi klasa ISBD:Resource.

Treba reći da su izričito deduktivni ili izričito induktivni pristup krajnje opcije u oblikovanju ontologije. Cilj i prve i druge metode određeni je ideal koji je nedostizhan. Deduktivni pristup teži dosljednom logičkom strukturiranju činjenica o svijetu što se zbog ograničenosti logičkog modeliranja kao takvog pokazuje neizvedivim: nisu svi odnosi među stvarima logički odnosi. U temelju je deduktivnog ili *top-down* pristupa, dakle, jedinstvena i univerzalna teorija svijeta, ili barem jednog njegova dijela, što je prilično nerealan cilj i u društvenom smislu. Naime, kao što nema teorije koja će interpretirati sve činjenice, isto tako, nema ni teorije koju će svi prihvatiti, koliko god dobro interpretirala činjenice. Prema tome, određena kontradiktornost i nedorečenost našeg znanja i naših teorija neizbježna je. S druge strane, induktivni ili *bottom-up* pristup teži iscrpnom i preciznom popisivanju i opisivanju svih dostupnih činjenica o određenom dijelu svijeta što se zbog njihova mnoštva i promjenjivosti pokazuje nepraktičnim i nedostiznim, a zbog fragmentarnosti instrumenata našeg opažanja i prikupljanja podataka – nesustavnim. Prema tome, možemo reći da su određena pojednostavljena i generalizacije nužni u našem znanju i tumačenju svijeta. Stoga OAP ontologija kombinira ova dva pristupa i pokušava postići određenu ravnotežu između:

- a) jasnoće i konzistentnosti u definiranju bibliografskih entiteta i njihovih semantičkih odnosa
- b) preciznosti i količine informacija koje opisuju te entitete u skupu autoriziranih podataka.

MADS/RDF i OAP imaju slično sintaktičko rješenje kada je u pitanju modeliranje pristupnica ili leksičkih oznaka. OAP modelira pristupnice i druge strukturirane nizove podataka isključivo kao RDF kolekcije, odnosno kao strukturu `rdf:List`. MADS/RDF, pored strukture `rdf:List`, dopušta i labaviju strukturu `rdf:Seq` koja pripada grupi tzv. RDF kontejnera. Bitna je razlika između MADS/RDF i OAP ontologije u granulaciji, odnosno razini specifičnosti ili detalja ontoloških elemenata koji služe za opisivanje entiteta u skupu autoriziranih podataka.

To je činjenica koja je razvidna iz Tablice 1: vidimo da OAP rječnik ima dvostruko veći broj elemenata.

**TABLICA 1.** Usporedba broja elemenata MADS/RDF, OAP i ADM ontološkog rječnika

Usporedba broja elemenata		
Ontologija	Klase	Svojstva
MADS/RDF	58	72
OAP	102	147
ADM	1	16

Za takav odnos broja elemenata postoje dva razloga:

1. MADS/RDF polazi od autoriziranih podataka koji su pohranjeni u zapisima formata MARC21 koji ima manji broj i jednostavniju strukturu podatkovnih elemenata od trećeg izdanja formata UNIMARC/A koje je prošireno znatnim brojem elemenata kako bi se izrazili prostorno-vremenski podaci o staroj građi, tiskarima i nakladnicima te kako bi se eksplicitno izrazili FRBR entiteti *djelo* i *izraz* kao aspekti naslova.
2. MADS/RDF namjerno zadržava općenitiju razinu modeliranja podataka kako bi bio prikladan za predstavljanje zajedničkog korpusa autoriziranih podataka koji upotrebljava najširi krug knjižnične zajednice. Proširivanje MADS/RDF ontologije dodatnim skupom specifičnijih elemenata prepušta se lokalnim aplikacijama. Nasuprot tomu, OAP ontologija modelira skoro sve semantičke elemente koji su definirani u trećem izdanju formata UNIMARC/A; izostavljeni su podatkovni elementi koji se smatraju redundantnima ili nefunkcionalnima u RDF podatkovnom modelu.

Naravno, ono što se tvrdi za MADS/RDF vrijedi i za OAP i svaku drugu ontologiju: RDF ontologije otvoreni su sustavi pa se klase i svojstva mogu isključivati iz upotrebe, odnosno dodavati i specijalizirati, kako bi se prilagodili potrebama korisnika podataka. Zajednica koja želi modelirati svoje metapodatke u RDF podatkovnom modelu mora odlučiti koja je strategija za nju najprikladnija: prihvaćanje postojećih rječnika i njihovo prilagođavanje ili izgradnja novih.

## LITERATURA

- DCMI Terms. (2012). *DCMI Metadata Terms*. 2012-06-14, <http://www.dublincore.org/documents/dcmi-terms/> (23-01-2019).
- DUNSIRE, G., WILLER, M. i PEROŽIĆ, P. (2013). Representation of the UNIMARC bibliographic data format in Resource Description Framework. *International Conference on Dublin Core and Metadata Applications, North America, Sep. 2013*, <http://dcpapers.dublincore.org/pubs/article/view/3680> (23-01-2019).
- FOAF. (2014). *FOAF Vocabulary Specification 0.99*, Namespace Document 14 January 2014 – *Paddington Edition*, <http://xmlns.com/foaf/spec/> (23-01-2019).
- Internationalized Resource Identifiers (IRIs), <http://www.w3.org/International/O-URL-and-ident.html> (23-01-2019).
- Library Linked Data Incubator Group Final Report. (2011). *Library Linked Data Incubator Group Final Report: W3C Incubator Group Report 25 October*, <http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/> (23-01-2019).
- MADS/RDF. (2015) *Namespace Document MADS/RDF (Metadata Authority Description Schema in RDF)*. V. 1.3.0, <http://www.loc.gov/standards/mads/rdf/v1.html> (23-01-2019).
- Open Metadata Registry* (OMR), <http://metadataregistry.org/> (23-01-2019).
- OWL Primer. (2012). *OWL 2 Web Ontology Language Primer*. 2nd ed. W3C Recommendation 11 December 2012. P. Hitzler, et al. (ur.), <http://www.w3.org/TR/2012/REC-owl2-primer-20121211/> (23-01-2019).
- RIVA, P., LE BŒUF, P. i ŽUMER, M. *IFLA Library Reference Model: a conceptual model for bibliographic information*, [https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017\\_rev201712.pdf](https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf) (23-01-2019).
- RIVA, P. i ŽUMER, M. (2015). Introducing the FRBR Library Reference Model. *IFLA WLIC 2015 Cape Town*, <http://library.ifla.org/1084/1/207-riva-en.pdf> (23-01-2019).
- PEROŽIĆ, P. (2012). Bibliografski podatci na webu podataka ili od UNIMARC zapisa do RDF grafa. U Hasenay, D. i Krtalić, M. (ur.) *15. Seminar Arhivi, knjižnice muzeji: mogućnosti suradnje u okruženju globalne informacijske infrastrukture: zbornik radova* (str. 101–135). Zagreb: Hrvatsko knjižničarsko društvo.
- PEROŽIĆ, P. (2016). *Modeliranje autoriziranih podataka u Općem okviru za opisanje resursa: doktorska disertacija*. Zadar: Sveučilište u Zadru.
- RDF 1.1 Primer. (2014). *RDF 1.1 Primer: W3C Working Group Note 24 June 2014*, <http://www.w3.org/TR/rdf11-primer/> (23-02-2019).
- RDF Primer. (2004). *RDF Primer: W3C Recommendation, 10 February 2004*. F.

- Manola i E. Miller (ur.), <http://www.w3.org/TR/rdf-primer/> (23-01-2019).
- SKOS Primer. (2009). *SKOS Simple Knowledge Organization System Primer: W3C Working Group Note, 18 August 2009*. A. Isaac i E. Summers (ur.), <http://www.w3.org/TR/skos-primer/> (23-01-2019).
- SKOS System Reference. (2009). *SKOS Simple Knowledge Organization System Reference: W3C Recommendation, 18 August 2009*. A. Miles i S. Bechofer (ur.), <http://www.w3.org/TR/skos-reference/> (23-01-2019).
- UNIMARC Authorities. (2009). *UNIMARC manual: authorities format*. 3rd ed. M. Willer (ur.). München: K. G. Saur.
- WILLER, M. i DUNSIRE, G. (2013). *Bibliographic information organization in the semantic web*. Oxford: Chandos.
- WILLER, M., DUNSIRE, G. i Perožić, P. (2013). The UNIMARC in RDF project: namespaces and linked data. *IFLA WLICK 2013 Singapore*, <http://library.ifla.org/156/1/222-willer-en.pdf> (23-01-2019).
- W3C RDF Validation Service*, <http://www.w3.org/RDF/Validator/> (23-01-2019).

# MODELLING AUTHORIZED METADATA IN RESOURCE DESCRIPTION FRAMEWORK

## KEYWORDS:

*Authority Data Ontology (Ontologija autoriziranih podataka – OAP), IFLA UNIMARC Authorities Format, authority data modelling, RDF, SKOS, SKOSXL*

## ABSTRACT

*The research addressed in this paper investigated the most suitable options, using the Resource Description Framework (RDF) data model, for representing authorized bibliographic metadata stored in IFLA UNIMARC Authorities Format (UNIMARC/A) as semantically related and machine understandable data. The outcome of the research is a concrete product, Authority Data Ontology (Ontologija autoriziranih podataka – OAP). OAP is based on the abstract syntax of the SKOSXL/RDF data model, and semantics that were modelled according to Format UNIMARC/A-specific logical data elements describing bibliographic entities. Since the OAP vocabulary is a simpler tool than the Format UNIMARC/A but transmits the same information it can be said that this approach achieves a kind of optimization.*

*In terms of methodology, the OAP ontology combines both bottom-up and top-down approaches in trying to strike a balance between: a) clarity and consistency in defining bibliographic entities as OAP classes and their semantic relations as OAP properties, and b) the accuracy and amount of information that describes these entities in the authorized data set.*