

# ОСНОВНЕ ОДЛИКЕ ОТВОРЕНИХ ПОВЕЗАНИХ ПОДАТАКА У БИБЛИОТЕКАМА

*Ерик Т. Мичел*

*Универзитет Калифорнија,*

*Сан Дијеџо, САД*

УДК: 004.738.5:02

004.65:02

001.103

## Сажетак

У другом поглављу *Извештаја о библиотечким технологијама*, под називом „Библиотечки повезани подаци: истраживање и усвајање“<sup>1</sup>, анализираћемо отворене повезане податке и отворене повезане речнике узимајући у обзир пет основних одлика метаподатака (модел података, правила о садржају, шема метаподатака, серијализација података и размена података). Ово поглавље пружа основу за разумевање технологије која се користи у студијама случаја у трећем поглављу и пружа преглед предности и недостатака у вези са отвореним повезаним подацима и отвореним повезаним речницима у библиотекама, архивима и музејима.

---

<sup>1</sup> Erik T. Mitchell, “Library Linked Data: Research and Adoption”, *Library Technology Reports* 49, no. 5 (July 2013).

Узимајући у обзир све проблеме, трендове и области истраживања о којима смо говорили у првом поглављу, дошли смо до закључка да је кључни фактор за развој квалитетних система метаподатака успостављање интероперабилности метаподатака без губитка података и усклађивање различитих домена.<sup>2</sup> Ово се може видети у спецификацијама за RDA формат, који је креиран с циљем да се користи за широк спектар грађе у различитим институцијама, као и у истраживањима која се баве питањем како се различити стандарди, као што су RDA, DACS и CCO међусобно уклапају и шта се може учинити да би се побољшала компатибилност различитих формата за каталогизацију, као и за садржај уопште.<sup>3</sup>

У широј веб-заједници постављају се слична питања о томе како најбоље креирати, делити, прикупљати и користити податке у различитим окружењима на вебу. Заједница библиотека, архива и музеја која се залаже за отворене повезане податке (LODLAM)<sup>4</sup> и група инкубатора за библиотечке повезане податке W3C конзорцијума имају улогу у прикупљању података и истраживању проблема у вези са семантичким и повезаним подацима за потребе институција које чувају културно наслеђе. Ове заједнице настале су услед потребе библиотека, архива и музеја за праћењем развоја технологије, начина креирања података и праксе објављивања података заједница окренутих вебу.

LODLAM:

<http://lodlam.net>

W3C Library Linked Data група инкубатора:

[www.w3.org/2005/Incubator/llid](http://www.w3.org/2005/Incubator/llid)

Принцип отворених повезаних података и отворених повезаних речника примењен је у бројним пројектима библиотека, архива и музеја, укључујући пројекат Сервис повезаних нормативних података и речни-

2 Craig Willis, Jane Greenberg, Hollie White, "Analysis and Synthesis of Metadata Goals for Scientific Data," *Journal of the American Society for Information Science and Technology* 63, no. 8 (August 2012): 1505–1520, doi:10.1002/asi.22683.

3 Joan C. Biella, Heidi G. Lerner, "The RDA Test and Hebraica Cataloging: Applying RDA in One Cataloging Community," *Cataloging and Classification Quarterly* 49, no. 7–8 (2011): 676–695, doi:10.1080/01639374.2011.616450; William E. Landis, "Plays Well with Others: DACS and CCO as Interoperable Metadata Content Standards," *VRA Bulletin* 34, no. 1 (2007): 97–103; Ilana Toloff, "The Path toward Global Interoperability in Cataloging," *Information Technology and Libraries* 29, no. 1 (March 2010): 30–39; Yuji Tosaka and Jung-ran Park, "RDA: Resource Description and Access—A Survey of the Current State of the Art," *Journal of the American Society for Information Science and Technology* 64, no. 4 (April 2013): 651–662.

4 Linked Open Data in Libraries, Archives and Museums.

ка Конгресне библиотеке, Европеанине повезане податке, Шведски централни каталог, Националну библиотеку Немачке, Британску националну библиографију, Отворену библиотеку и нову Дигиталну јавну библиотеку Америке (DPLA). Док се опсег и резултати ових пројеката разликују, сваки од њих пружа добар пример како метаподаци библиотека, архива и музеја могу да мењају облик и сврху и како могу да се користе више пута. С тим у вези, постоје бројна истраживања о изгледу, употреби и последицама употребе принципа отворених повезаних података за податке институција које чувају културно наслеђе.<sup>5</sup>

Сервис повезаних нормативних података и речника Конгресне библиотеке:  
<http://id.loc.gov>

Еuropeанини отворени повезани подаци:  
<http://data.europeana.eu>

Централни каталог Шведске:  
[www.libris.kb.se](http://www.libris.kb.se)

Национална библиотека Немачке:  
[www.dnb.de](http://www.dnb.de)

Национална библиографија Британије:  
[www.bl.uk/bibliographic/natbib.html](http://www.bl.uk/bibliographic/natbib.html)

Отворена библиотека:  
<http://openlibrary.org>

Дигитална јавна библиотека Америке (DPLA):  
<http://dp.la>

Како бисмо разјаснили како ови системи функционишу и каква је улога отворених повезаних података и отворених повезаних речника, у овом поглављу детаљно ћемо анализирати отворене повезане податке узимајући у обзир пет основних одлика метаподатака, представљених у првом поглављу. Пре него што приступимо анализи отворених повезаних података, пружићемо кратак преглед сваке од пет основних одлика метаподатака.

---

5 Getaneh Alemu, Brett Stevens, Penny Ross and Jane Chandler, "Linked Data for Libraries: Benefits of a Conceptual Shift from Library-Specific Record Structures to RDF-Based Data Models" (conference paper, IFLA Council and General Conference, Helsinki, Finland, August 11–17, 2012); Bernhard Haslhofer, Elaheh Momeni, Bernhard Schandl and Stefan Zander, "Europeana RDF Store Report," Europeana Connect, Results and Resources, March 8, 2011, [www.europeanaconnect.eu/documents/europeana\\_ts\\_report.pdf](http://www.europeanaconnect.eu/documents/europeana_ts_report.pdf).

## ШТА СУ ОТВОРЕНИ ПОВЕЗАНИ ПОДАЦИ?

Принцип отворених повезаних података је широко заступљен и заснива се на раду Семантичког веба. Детаљно га је описао Тим Бернерс Ли.<sup>6</sup> Овај принцип састоји се од две основне идеје. Прва је да подаци који се објаве на вебу могу да се лако повезују са већ постојећим сродним подацима и притом треба да буду разумљиви, како људима, тако и рачунарима. Друга идеја од кључног значаја за овај принцип је да подаци морају бити отворени и ослобођени правних ограничења у вези са ауторским правима, како би могли да се повезују и вишеструко користе. Отворени повезани речници засновани су на истом принципу као и отворени повезани подаци, с тим што се код отворених повезаних речника не објављују подаци (ресурси и информације о ресурсима), већ речници који доприносе изградњи семантичког веба (листе термина, шеме метаподатака, таксономије и онтологије).

Отворени повезани подаци револуционализоваће веб тако што ће омогућити изградњу информационе мреже која ће бити разумљива и човеку и рачунару, због чега ће бити употребљивија од података које може да разуме само човек. Предност отворених повезаних података је у томе што омогућавају размену података користећи заједничке концепте и структуре података који не подлежу ауторским правима. Ови речници обухватају рудиментарне речнике (нпр. тип ресурса), као и комплексне онтологије (нпр. „пријатељ пријатеља“) које помажу да се неки ресурс смести у широкој бази информација. Из тог разлога, разумљиво је што су се отворени повезани подаци наметнули као систем за размену података између библиотека, а отворени повезани речници представљају нит која повезује те податке. Постоје бројне иницијативе у оквиру заједнице библиотека, музеја и архива да се отворени повезани речници објављују, укључујући и Регистар отворених метаподатака, као и Сервис повезаних нормативних података и речника.

Регистар отворених метаподатака:  
<http://metadataregistry.org>

С обзиром на то да постоје бројне публикације у којима су набројани ти речници и које говоре о алатима и развијеним веб-сервисима, водећи се принципом отворених повезаних података и речника,<sup>7</sup> нећемо их овде

---

<sup>6</sup> Tim Berners-Lee, "Linked Data," W3C website, last updated June 18, 2009, [www.w3.org/DesignIssues/LinkedData.html](http://www.w3.org/DesignIssues/LinkedData.html).

<sup>7</sup> Karen Coyle, "Linked Data Tools: Connecting on the Web," *Library Technology Reports* 48, no. 4 (May/June 2012); Bernhard Haslhofer and Antoine Isaac, "data.europeana.eu: The European Linked Open Data Pilot," in *Metadata Harmonization: Bridging Languages of Description*: September 21–23, 2011, the Hague, Netherlands, ed. Thomas Baker, Diane I. Hillmann and An-

детаљно набрајати. Уместо тога, бавићемо се питањем како отворени повезани подаци и речници могу да послуже као нови модел за метаподатке библиотека, архива и музеја, тако што ћемо размотрити како се они уклапају у пет основних одлика метаподатака (модел података, правила о садржају, шема метаподатака, серијализација података и размена података). Како бисмо боље разумели како се речници уклапају у структуру повезаних података, отворене повезане речнике посматраћемо као саставни део отворених повезаних података. Зато ћемо се приликом анализе усредсредити на отворене повезане податке, а речнике ћемо сматрати саставним делом шеме метаподатака. Упркос овом поједностављењу, битно је истаћи да су сами отворени повезани речници креирани уз помоћ спецификација за отворене повезане податке. Притом, иако је веома битно да су подаци отворени, то је питање ауторских права и лиценцирања и нема везе са метаподацима, па због тога није кључно за наше истраживање.

Укратко, повезани подаци заснивају се на четири основна принципа (преузето и прилагођено од Бернерс Лија): 1) коришћењу јединственог идентификатора извора (URI), 2) коришћењу протокола за пренос хипертекста јединственог идентификатора извора (HTTP URI), 3) добијању употребљивих података када се HTTP URI разреши, 4) укључивању линкова ка другим јединственим идентификаторима извора. Бернерс Ли користи и систем рангирања са пет звездица, где једна звезда значи да подаци на вебу имају отворену лиценцу, а пет звездица значи да су подаци у потпуности повезани, отворени и машински читљиви у облику RDF докумената.<sup>8</sup>

Бернерс Ли је током ТЕД конференције говорио о отвореним повезаним подацима као кључном кораку за изградњу семантичког веба.<sup>9</sup> Представио је окружење у коме рачунари и људи заједнички идентификују објекте, не само уз помоћ текста, већ и коришћењем дефинисаних речника и линкова ка ресурсима на вебу, што је илустровао једноставним примером тако што је на виртуелној платформи за мапирање обележио салу у којој је држао презентацију. Тај једноставни допринос, како је истакао, могу убудуће вишеструко користити и људи и рачунари. Ова релативно једноставна замисао представља суштину принципа отворених повезаних података и основу за системе метаподатака библиотека, архива и музеја који желе да допринесу овој идеји, не тако што ће пружити текстуалне записе о ресурсима, већ тако што ће помоћи да се ресурси из више база података повежу темељно и прецизно.

---

toine Isaac (Dublin, OH: Dublin Core Metadata Initiative, 2011), 94–104; Library of Congress, "The BIBFRAME Model: Vocabulary Updates," Bibliographic Framework Initiative website, accessed April 20, 2013, <http://bibframe.org/vocab>.

<sup>8</sup> Berners-Lee, "Linked Data".

<sup>9</sup> Tim Berners-Lee, "Tim Berners-Lee on the Next Web", TED video, 16:17, filmed February 2009, posted March 2009, [www.ted.com/talks/tim\\_berniers\\_lee\\_on\\_the\\_next\\_web.html](http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html).

Како бисмо их боље разјаснили, у наставку овог поглавља приступићемо анализи основних одлика отворених повезаних података.

## ПРВА ОДЛИКА: МОДЕЛ ПОДАТАКА ЗА ОТворЕНЕ ПОВЕЗАНЕ ПОДАТКЕ

Модел података представља основну структуру метаподатака која одређује изглед једног записа. Код RDF<sup>10</sup> модела, којим ћемо се овде бавити, користи се модел података са чворовима и луковима (граф). Насупрот томе, код MARC записа користи се раван модел података код кога су метаподаци везани за један запис. Остали модели података, које нећемо разматрати у овом поглављу, су модели објекти-везе и модели засновани на објектима.

За отворене повезане податке уобичајено се користи RDF модел података. RDF представља модел путем кога се могу обухватити метаподаци о некој грађи, али разликује се од осталих стандарда, као што су MARC, EAD и други модели који се заснивају на записима, по томе што се усредсређује на појединачне вредности ресурса (нпр. наслов књиге је *Аваншуре Хаклбе-ри Фина*). Сам RDF је само модел података који представља средство за формирање простих записа и њихово повезивање у низ који представља потпун дескриптивни запис за одређени ресурс. W3C конзорцијум је 1999. године предложио RDF као модел за отворене податке, а 2004. године објављене су измењене и допуњене спецификације.<sup>11</sup> RDF представља основу семантичког веба, као и технологија попут OWL, SKOS и повезаних података.

Потпуно разумевање RDF модела захтева много више простора од онога који нам је овде на располагању, али постоје бројни одлични приручници за RDF модел. Неки од њих су W3C RDF документација (почевши од W3C водича кроз основе RDF модела) и W3C списак литературе о семантичком вебу. Поред тога, корисно је и друго поглавље Лианг Јуове књиге „Водич за развој семантичког веба“, које пружа једноставан увод у RDF модел, корак по корак кроз примере, почевши од најједноставнијих ка сложенијим.<sup>12</sup> С обзиром на приступачност ове методе, наше истраживање RDF модела из перспективе библиотека, архива и музеја заснива се на Јуовим објашњењима.

W3C водич кроз основе RDF модела (са линковима ка осталој документацији за RDF модел):

---

<sup>10</sup> Resource Description Framework.

<sup>11</sup> Frank Manola and Eric Miller, eds., “RDF Primer,” W3C Recommendation, February 10, 2004, [www.w3.org/TR/rdf-primer](http://www.w3.org/TR/rdf-primer).

<sup>12</sup> Liyang Yu, “The Building Block for the Semantic Web: RDF,” chapter 2 in *A Developer’s Guide to the Semantic Web* (Heidelberg; New York: Springer, 2011), 19–86.

[www.w3.org/TR/2004/REC-rdf-primer-20040210](http://www.w3.org/TR/2004/REC-rdf-primer-20040210)

W3C: литература о семантичком вебу:

[www.w3.org/2001/sw/wiki/Books](http://www.w3.org/2001/sw/wiki/Books)



### *Шема 2.1*

RDF исказ састоји се од субјекта, предиката и објекта

Једна од кључних предности RDF модела је то што може да опише физичке објекте, дигиталне објекте, као и везу између примарног објекта и његовог сурогата. То је било од пресудног значаја за његов одабир као стандарда за библиотеке, архиве и музеје и из тог разлога системи који су у центру развоја метаподатака за 2012. годину користе RDF модел као основу.

RDF модел састоји се од исказа, који се још називају и RDF тројке. Један исказ састоји се од субјекта, предиката и објекта и има структуру чворова и лукова графа. У Шеми 2.1 приказана је основна структура исказа. Треба истаћи да је исказ усмерен, у смислу да је предикат назив за одређену везу која спаја субјекат и објекат. Оваква синтакса је у суштини веома слична моделу поља, потпоља и вредности код MARC формата и моделу елемента, атрибута и вредности који се користи у моделима метаподатака који се заснивају на запису. Из тог разлога, предикати, који се другачије називају својства, веома су слични елементима метаподатака код других модела.

RDF тројка састоји се од два чвора (субјекат и објекат) и лука (предикат) који представља везу између ресурса и субјекта. Сматра се да је RDF модел података флексибилнији од модела који се заснивају на запису, као што је MARC, због тога што је сваки исказ потпун и не зависи од других исказа. Насупрот томе, свако поље у MARC запису зависи од контекста у ширем запису и без тог контекста не би значило пуно. Како би се постигла таква флексибилност у тако једноставном моделу (субјекат–предикат–објекат), RDF модел дозвољава објектима и предикатима да сами постану субјекти, што ствара разгранату структуру која може да се неограничено шири и није неодређена и ограничена, што може да буде случај са равним

моделима података као што је MARC. То је предност RDF база података, које се другачије називају граф базе, насупротив релационим базама података које захтевају унапред дефинисану шему и имају ограничење у погледу веза које могу да се остваре.

Увођење модела који се заснивају на графу и дрвету (нпр. FRBR) у систем библиотечке каталогизације довео је до тога да MARC постане неодговарајући модел за чување и серијализацију података, зато што структура MARC модела, заснована на запису, захтева значајно дуплирање метаподатака и ослања се на слободан текст, уместо на јединствене идентификаторе извора (URI). Овај недостатак MARC модела (немогућност да лако покаже отворене везе) представља највећу предност RDF модела која се огледа, како у обликовању података, тако и у могућности за проширење система. Из тог разлога се за RDF каже да дозвољава неограничене везе. То заправо значи да предикати и објекти могу да сами постану ресурси са сопственим предикатима и објектима и да на овај начин један ресурс може прецизно да прати сложене односе са другим, екстерним ресурсима.

Табела 2.1

Пример синтаксе извучене из MARC записа за књигу *Аваншуре Хаклбери Фина*

The Adventures of Huckleberry Finn	is_a	"Text"
The Adventures of Huckleberry Finn	is_written_by	"Mark Twain"
The Adventures of Huckleberry Finn	has_publication_date	"1884"
The Adventures of Huckleberry Finn	has_length	"438 pages"
The Adventures of Huckleberry Finn	is_published_by	"Chatto & Windus"
The Adventures of Huckleberry Finn	is_about	"Finn, Huckleberry (Fictitious Character) Fiction"
The Adventures of Huckleberry Finn	is_about	"Runaway children, Fiction"

У наставку анализе основних елемената у овом поглављу, користићемо пример књиге *Аваншуре Хаклбери Фина*<sup>13</sup> Марка Твена, коју је први пут издала штампарија Чато и Виндус 1884. године. Запис у MARC формату за ову књигу може се пронаћи у виртуелном каталогу Конгресне библиотеке.

<sup>13</sup> The Adventures of Huckleberry Finn.



Извлачењем метаподатака из овог записа и њиховим пребацивањем у RDF формат показаћемо предности модела података са чворовима и луковима и употребе јединствених идентификатора (URI) уместо вредности слободног текста. Неколико таквих исказа приказано је у Табели 2.1, укључујући исказ о аутору, издавачу и тематским одредницама. Овај први корак у екстракцији метаподатака и њиховом представљању у виду исказа графа путем табеле не разликује се пуно од MARC формата. Притом, у Табели 2.1 и Шеми 2.2 користе се дословне вредности уместо URI идентификатора, тако да је у Шеми 2.2 само делимично примењен RDF модел. У примерима који следе показаћемо како се та библиографска структура података може унапредити и укључићемо више метаподатака у RDF формат.

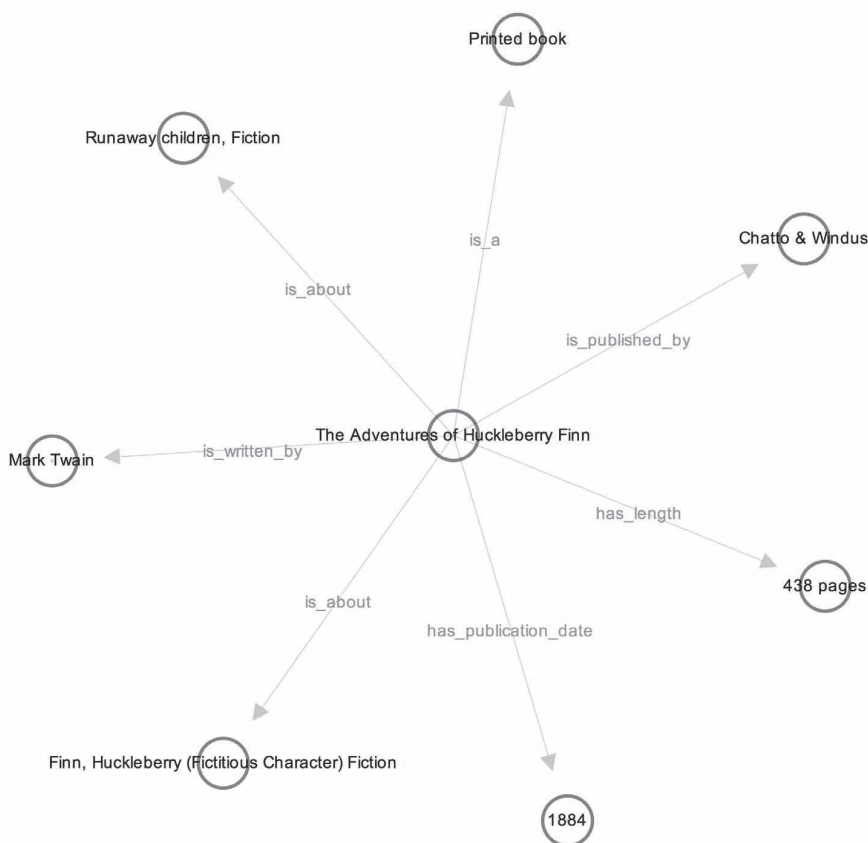
MARC запис за *Аваншуре Хаклбери Фина*:  
<http://lcn.loc.gov/35020965>

У Шеми 2.2 представићемо садржај табеле у виду графа. Иако су предикати формирани донекле слободно, они су у грубој вези са пољима за аутора, издавача, датум издавања и тематску одредницу MARC формата. Овај модел је у великој мери прецизан, али, с обзиром да нисмо превели дословне вредности у URI идентификаторе, принуђени смо да користимо наслов књиге као примарни идентификатор. Тиме се ослањамо на концепт „главне каталожке јединице“ из традиционалне каталожке праксе, што није практично уколико имамо наслов који има пуно издања, као у нашем примеру.

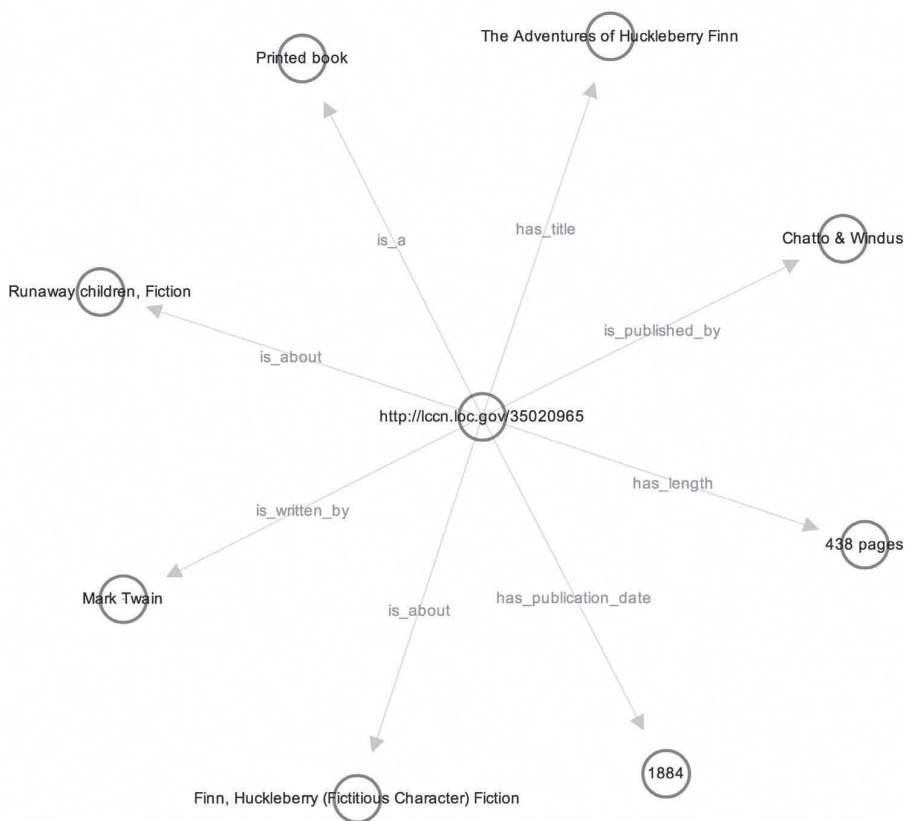
Како бисмо решили овај проблем, одабраћемо јединствени идентификатор који ће упућивати на локацију где постоји више информација о наслову. Тај идентификатор засад може бити јединствена локација извора (URL) за запис који се налази у каталогу Конгресне библиотеке. То можемо видети у Шеми 2.3. URL је искоришћен као јединствени идентификатор извора (URI) за дати ресурс. Наслов је постао објекат у новој синтакси са предикатом *has title*. Ова једноставна промена је значајна јер нам омогућава да све исказе о овом делу спојимо једним URI идентификатором.

Како смо заменили идентификатор ресурса URI идентификатором, пожељно је и да одредимо идентификаторе за остале дескриптивне метаподатке. Како ћемо видети у наредним поглављима, постоји све више извора URI идентификатора. У неким случајевима, одговараће нам веза ка неком екстерном речнику (нпр. за тип ресурса). У неким другим случајевима, можда ће нам одговарати веза ка неким другим ресурсима (нпр. за сродно дело, аутора или предметну одредницу). Наш једноставни скуп исказа у Табели 2.1 укључује предметну упутницу, именску упутницу, временску упутницу и још неке вредности које можда немају дефинисан речник. Могу се користити дословне вредности уколико је то потребно. За-

право, у одређеном тренутку оне постају неопходне, како би садржај записа био разумљив и човеку. Зарад једноставности, користимо Сервис повезаних нормативних података и речника Конгресне библиотеке за URI идентификаторе. Када заменимо дословне вредности нормативним именима и предметним ознакама, податке је лакше повезати. То је приказано у Табели 2.2.



Шема 2.2  
RDF искази о књизи *Аваншуре Хаклбери Фина*



Шема 2.3

RDF искази о књизи *Аваншуре Хаклбери Фина* са пермалинком Конгресне библиотеке као URI идентификатором

У Табели 2.2 пружен је једноставан преглед ових података где су вредности за субјекат и објекат замењене URI идентификаторима. Засад нећемо конвертовати друге вредности, као што су „Printed book” („штампана књига”) у URI иако постоје отворени повезани речници који би олакшали пребацивање ових вредности у низ тројки које су у складу са отвореним повезаним речницима. Иако још увек нисмо изменили предикате, већ можемо да видимо како би URI идентификатори помогли у прикупљању исказа за дато дело, без икаквих двосмислености, док би URI идентификатори за предметне одреднице, аутора и друге субјекте помогли рачунарима да спроведу детаљну унакрсну претрагу и поступке за проширење упита.

Табела 2.2

Искази, измењени тако да URI идентификатори замењују дословне вредности извучене из MARC записа

http://lccn.loc.gov/35020965	is_a	"Printed book"
http://lccn.loc.gov/35020965	is_written_by	http://id.loc.gov/authorities/names/n79021164
http://lccn.loc.gov/35020965	has_publication_date	"1884"
http://lccn.loc.gov/35020965	has_length	"438 pages"
http://lccn.loc.gov/35020965	is_published_by	http://id.loc.gov/authorities/names/n85242407
http://lccn.loc.gov/35020965	is_about	http://id.loc.gov/authorities/subjects/sh2008103799
http://lccn.loc.gov/35020965	is_about	http://id.loc.gov/authorities/subjects/sh2008110345

Табела 2.3

Предикати мапирани према Dublin Core екстерном вокабулару

Име предиката	Име из екстерног вокабулара	Дословни објекат или URI
is_a	dc:type	"text"
is_written_by	dc:creator	http://id.loc.gov/authorities/names/n79021164
has_publication_date	dc:date	"1884"
has_length	dc:extent	"438 pages"
is_published_by	dc:publisher	http://id.loc.gov/authorities/names/n85242407
is_about	dc:subject	http://id.loc.gov/authorities/subjects/sh2008103799
is_about	dc:subject	http://id.loc.gov/authorities/subjects/sh2008110345
title	dc:title	"The Adventures of Huckleberry Finn"

Када вредности субјекта и предиката заменимо URI идентификаторима, можемо да користимо URI идентификаторе и за одређивање предиката. То има бројне предности, укључујући и прецизно усклађивање предиката са речницима и шемама метаподатака. Апсолутним идентификовањем предиката отклања се могућност за постојање двосмислености, што повећава интероперабилност и поузданост аутоматске обраде података. На пример, тај ниво поузданости омогућава откривање ненаглашених веза између извора. У најбољим случајевима, то доводи и до вишеструког коришћења заједничких вокабулара за URI идентификаторе предиката.

Док се код метаподатака за отворене повезане податке најчешће користи више вокабулара и шема метаподатака за опис, зарад једноставности, у нашем објашњењу RDF формата користимо само Dublin Core шему. У Табели 2.3 конструисана имена предиката мапирана су по елементима из Dublin Core речника.

Свако од ових екстерних имена вокабулара (нпр. dc:title) користи и простор имена и име елемента. Иако ћемо детаљније анализирати простор имена када будемо говорили о серијализацији RDF модела, вреди истаћи да је простор имена једна од саставних функција XML формата која омогућава скраћивање имена из других URI идентификатора, као нпр. <http://purl.org/dc/elements/1.1/title> у dc:title. То се постиже уз помоћ xmlns атрибута. Пример за примену овога можете погледати у Приказу 2.1.

## АПСТРАКТНИ RDF МОДЕЛ

У овој сажетој анализи RDF модела представили смо RDF тројке (Шема 2.1) које су основа RDF синтаксе и анализирали смо исказе који се заснивају на URI идентификаторима који олакшавају стварање описа сложених односа уз помоћ екстерних стандарда и речника. Подразумева се да се најбољи водичи за коришћење RDF формата придржавају уобичајене библиотечке праксе, укључујући и предности коришћења нормативних података, потребу да се стандарди користе вишеструко када је то могуће, и потребу да се садржај опише детаљно и недвосмислено.

Поредећи MARC и RDF модел података, дошли смо до закључка да RDF модел омогућује сложенији приступ опису када говоримо о комплексним односима где је један елемент у вези са више других елемената. Као што смо истакли на почетку излагања о RDF моделу, он представља основу, како семантичког веба, тако и отворених повезаних података. Како бисмо разумели улогу других елемената, као што су речници и крајње тачке повезаних података, прво ћемо се посветити анализи начина на који се пише RDF шема, другим речима серијализацији. Када говоримо о начинима серијализације RDF модела, присетимо се основних правила које је истакао

Ју.<sup>14</sup> RDF оптимизује информације за машинску обраду. За RDF могу да се користе и експлицитни и имплицитни искази, а с обзиром на то да се у RDF моделу креирају метаподаци уз помоћ недвосмислених URI идентификатора, могуће је прикупљати исказе о неком ресурсу из бројних извора. Иако нам модел података донекле помаже у овоме, заправо, структура и системи који је тумаче омогућавају спровођење ових циљева.

## ДРУГА ОДЛИКА: ПРАВИЛА О САДРЖАЈУ

Правила о садржају су важан део шема библиотечких метаподатака. Стварни примери правила о садржају подразумевају RDA и AACR2 и они уопштено одређују правила по којима се подаци извлаче из каталошког ресурса. Правила о садржају у блиској су вези са шемама метаподатака (нпр. када се извуче наслов ресурса, он се чува у пољу за наслов које је у вези са шемом метаподатака), али представљају засебне ентитете.

Принципи за креирање садржаја за отворене повезане податке користећи RDF модел су донекле флексибилни. Док RDF модел има строге смернице о томе како се дефинишу искази и како се успостављају односи између њих, он не одређује строго садржај који се уноси у RDF исказе. Заправо, ово је предност RDF модела података. Користећи структуру графа која се строго не ослања на унапред одређена својства, модел повезаних података може се лако проширити тако да укључује и друге дескриптивне податке.

С обзиром на то да RDF нема у себи правила за представљање дословних вредности, изузев једноставне синтаксе и команди за локализацију, ослања се на употребу екстерних речника и онтологија за садржај. То можете погледати у Приказу 2.1, где је искоришћена Библиографска онтологија како би се дефинисали типови ресурса који се каталогизују. Библиографска онтологија, позната и по простору имена BIBO, користи се у бројним пројектима за отворене повезане податке у библиотекама, архивима и музејима, укључујући и пројекат *Америчка хроника*<sup>15</sup> Конгресне библиотеке.

Библиографска онтологија:

<http://bibliontology.com>

Америчка хроника:

<http://chroniclingamerica.loc.gov>

---

<sup>14</sup> Liyang Yu, "The Building Block for the Semantic Web: RDF," chapter 2 in *A Developer's Guide to the Semantic Web* (Heidelberg; New York: Springer, 2011), 72.

<sup>15</sup> *Chronicling America*.

## ТРЕЋА ОДЛИКА: ШЕМА МЕТАПОДАТАКА У ОТВОРЕНИМ ПОВЕЗАНИМ ПОДАЦИМА

Шеме метаподатака су можда најразумљивија основна одлика метаподатака. Шеме су у вези са правилима која одређују структуру и садржај једног записа или једне тројке. Шеме метаподатака и правила о садржају блиско су повезани, али битно је разликовати их, с обзиром на њихову различиту сврху. На пример, док Dublin Core шема дефинише субјекат са смерницама за коришћење речника, поступак примене тог речника за одређени ресурс захтева правила о садржају која нису део дефиниције шеме метаподатака.

### RDF ШЕМА

Примери RDF модела помогли су нам да илуструјемо погодност овог модела података за представљање ресурса, али су нам и показали значај екстерних вокабулара за поступке описивања и представљања. Притом, RDF модел поседује и структуре које нам омогућавају да представимо сложеније односе између ресурса. Ове структуре део су RDF шеме (RDFS) која садржи речник који помаже у успостављању правила и структуре исказа који описују ресурсе.<sup>16</sup> То значи да је RDFS заправо језик за грађење речника, а не језик за опис. Коришћењем ових структура, можемо да побољшамо моћ рачунара да одреди везе између ресурса, помажући нам да од описа ресурса дођемо до приказа знања.

Као и код екстерних речника које смо користили у претходним примерима, и на RDFS речник може се упутити путем линка <http://www.w3.org/2000/01/rdf-schema#>. RDF шема састоји се од три основна типа структуре. То су класе, својства и корисничка својства. Кључна одлика RDF шеме је могућност да прикаже структуру родитељ–дете. У RDF шеми они су познати као надређене и подређене везе и односе се и на класе и својства (`rdfs:subClassOf`, `rdfs:subPropertyOf`). RDFS такође омогућује дефинисање типова података (`rdfs:Datatype`), домен (`rdfs:domain`) и скуп прихватљивих вредности (`rdfs:range`) за одређену класу или својство. И домен и скуп прихватљивих вредности имају софистицираније парњаке у Једноставном систему за организацију знања (SKOS)<sup>17</sup> и Језику онтологије на вебу (OWL)<sup>18</sup> – речницима који се баве везама вишег реда, укључујући прелазне и непрелазне односе и односе са и без преклапања. Поред односа родитељ–дете, типова података и дефиниције домена и скупа прихватљивих

---

<sup>16</sup> Dan Brickley and R. V. Guha, eds., "RDF Vocabulary Description Language 1.0: RDF Schema," W3C Recommendation, February 10, 2004, [www.w3.org/TR/rdf-schema](http://www.w3.org/TR/rdf-schema).

<sup>17</sup> Simple Knowledge Organization System.

<sup>18</sup> Web Ontology Language.

вредности, RDFS поседује и „погледај и“ својство (`rdfs:seeAlso`), својство за дефиницију (`rdfs:isDefinedBy`) и својство за обележавање (`rdfs:label`).

## ОСНОВЕ ЗА ГРАЂЕЊЕ РЕЧНИКА: SKOS И OWL

Метаподаци код којих је коришћен RDF испуњавају основни услов за отворене повезане податке, али уколико желимо да пратимо улогу неког ресурса у сложеном систему знања, морамо да користимо речнике који су развијени коришћењем напредне верзије RDF шеме. SKOS и OWL су такви системи који подржавају изградњу софистицираних речника (отворени повезани речници), који се такође називају таксономије и онтологије. SKOS и OWL подржавају интероперабилност дефинисањем односа између ресурса на начин који превазилази једноставну синтаксу са предикатом (нпр. `has_subject`). Ово је нарочито значајно уколико спајамо изворе података из различитих скупова података за отворене повезане податке, јер OWL својства попут `owl:sameAs`, `owl:differentFrom` и `owl:disjointWith` омогућавају аутоматско закључивање на основу прикупљених података.

Поред тога што SKOS и OWL речници омогућавају опис односа између ресурса код отворених повезаних података, они код отворених повезаних речника омогућавају сложене односе који су неопходни за миграцију сложених таксономија, класификационих шема и онтологија. Нормативна шема за опис метаподатака у RDF моделу (MADS/RDF)<sup>19</sup>, коју је осмислила Конгресна библиотека, представљена је као OWL онтологија. Не можемо пружити детаљан опис SKOS и OWL речника, али постоје бројни квалитетни приручници у вези са Protégé платформом који пружају увид у овај стандард кроз практичне примере.

Основни приручник за MADS/RDF Конгресне библиотеке:  
[www.loc.gov/standards/mads/rdf](http://www.loc.gov/standards/mads/rdf)

Protégé платформа:  
<http://protege.stanford.edu>

### ПРИМЕР ШЕМЕ ЗА МЕТАПОДАТКЕ БИБЛИОТЕКА, АРХИВА И МУЗЕЈА, ЗАСНОВАНЕ НА RDF МОДЕЛУ: ИНИЦИЈАТИВА ЗА ОТВОРЕНЕ АРХИВЕ И ВИШЕСТРУКУ УПОТРЕБУ И РАЗМЕНУ ОБЈЕКТА (OAI-ORE)

OAI-ORE<sup>20</sup> спецификација настала је из визије о омогућавању отвореног архивирања метаподатака уз решавање проблема који се могу

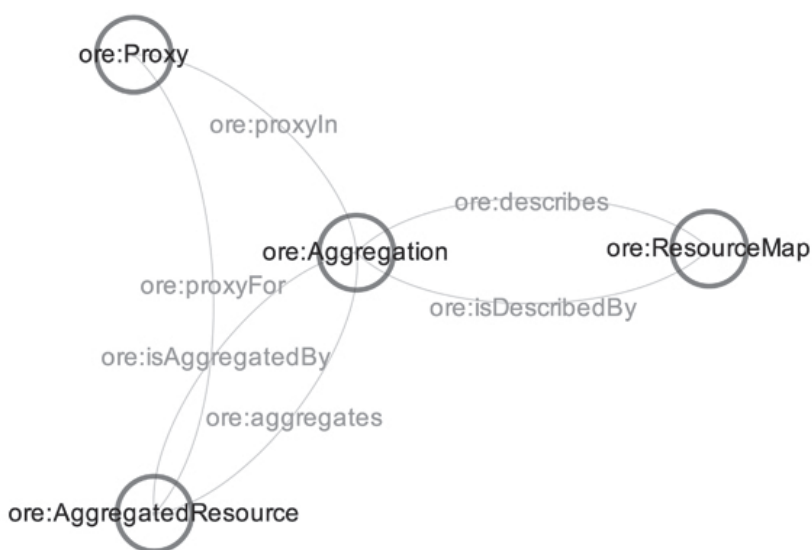
---

<sup>19</sup> The Metadata Authority Description Schema in RDF.

<sup>20</sup> Open Archives Initiative Object Reuse and Exchange.



јавити приликом прикупљања метаподатака из вишеструких извора.<sup>21</sup> Коришћењем Протокола за прикупљање метаподатака (OAI-PMH<sup>22</sup>), откривени су одређени изазови у вези са прикупљањем података из више извора, укључујући питања о пореклу, верзији, структури објеката и квалитету метаподатака. OAI заједница и даље активно ради на развоју ових спецификација, укључујући рад на Протоколу за прикупљање метаподатака (PMH), спецификацијама за вишеструку употребу и размену објеката (ORE) и новом пројекту осмишљеном тако да олакша синхронизацију ресурса кроз разне системе, под називом ResourceSync<sup>23</sup>. ORE спецификације се све више користе у системима дигиталних библиотека<sup>24</sup> и представљају основну компоненту нових спецификација за метаподатке, као што је Европeанин модел података (EDM)<sup>25</sup>.



#### Шема 2.4

Једноставан граф који приказује везу између четири основна ORE ентитета

<sup>21</sup> Herbert Van de Sompe and Carl Lagoze, "The Santa Fe Convention of the Open Archives Initiative," *DLib Magazine* 6, no. 2 (February 2000), doi:10.1045/february2000-vandesompe-oai.

<sup>22</sup> Protocol for Metadata Harvesting.

<sup>23</sup> Martin Klein, Robert Sanderson, Herbert Van de Sompe, Simeon Warner, Bernhard Haslhofer, Michael Nelson and Carl Lagoze, "ResourceSync Framework Specification—Beta Draft," February 1, 2013, [www.openarchives.org/rs/0.5/resourcesync](http://www.openarchives.org/rs/0.5/resourcesync).

<sup>24</sup> Michael Witt, "Object Reuse and Exchange (OAI/ORE)," *Library Technology Reports* 46, no. 4 (May/June 2010): 9.

<sup>25</sup> Antoine Isaac and Robina Clayphan, eds., "Europeana Data Model Primer: Europeana 1.0," Europeana Professional website, October 26, 2011, <http://pro.europeana.eu/documents/900548/770bdb58-c60e-4beb-a687-874639312ba5>.

Рад Мајкла Вита на OAI-ORE иницијативи у *Извештају о библиотечким технологијама* пружа детаљан увод у овај стандард, па га препоручујемо читаоцима који нису упућени у ову тему.<sup>26</sup> Ми ћемо у овом поглављу дати кратак преглед ORE спецификација, с обзиром на то да ћемо их користити у трећем поглављу када будемо говорили о EDM моделу.

ORE спецификације састоје се од четири основна ентитета: агрегације, ресурса агрегације, мапе ресурса и проксија. Основна сврха ORE онтологије је да створи модел за описивање ресурса са наглашеном разликом између примарног објекта и његових сурогата, као и између више верзија ресурса.

Ентитет за агрегацију (`ore:Aggregation`) прикупља ресурсе у облику директне везе са једном или више мапа ресурса, једним или више ресурса агрегације и једним или више проксија. Агрегација је највиши ентитет у једној ORE збирци. Мапе ресурса прикупљају ресурсе и структурне и дескриптивне исказе у вези са њима. Ови ентитети имају бидирекционална својства која показују везе између ентитета. На пример, агрегација је описана мапом ресурса (`ore:isDescribedBy`), а мапа ресурса описује агрегацију (`ore:describes`).

Исто тако, агрегацијом се прикупљају ресурси агрегације (`ore:aggregates`), а прикупљени ресурси могу се описати преко проксија (`ore:Proxy` и `ore:ProxyFor`). Код ORE спецификација, проксији могу да представљају ресурсе агрегације и првенствено се користе када није потребан контекст о додељеним вредностима. У Шеми 2.4 представљен је графички приказ основних веза између ових ентитета, као и својстава (предиката) који их повезују.

Везе које су представљене у Шеми 2.4 представљају основне везе у ORE моделу и не показују нпр. како се путем агрегације могу прикупити више-струки ресурси. За наше истраживање система отворених повезаних података довољно је да разумемо ова четири ентитета у ORE моделу и својства која их повезују.

## **ЧЕТВРТА ОДЛИКА: СЕРИЈАЛИЗАЦИЈА ОТВОРЕНИХ ПОВЕЗАНИХ ПОДАТАКА**

Серијализација је заједничко име за моделе кодирања за метаподатке библиотека, архива и музеја. Серијализација се односи на начин чувања ресурса на физичком или дигиталном медију за складиштење. Неки од уобичајених модела за серијализацију којим ћемо се позабавити су XML, JSON и N3. Серијализација је често у вези са моделом података, а у случају MARC формата, и у вези са шемом метаподатака.

---

<sup>26</sup> Witt, "Object Reuse and Exchange".

Примена RDF апстрактног модела захтева употребу серије елемената метаподатака који се код RDF модела називају RDF речник. Овај речник налази се на URI идентификатору <http://www.w3.org/1999/02/22-rdf-syntax-ns#> и обично је дефинисан префиксом за простор имена rdf: у RDF документима. RDF речник састоји се од четири групе елемената: синтаксе, класе, својстава и ресурса. У Табели 2.4 пружен је кратак преглед RDF елемената који могу бити корисни за примере.

Табела 2.4

Списак RDF типова елемената, имена елемената и њихове опште употребе

Тип RDF елемента	RDF елемент	Општа сврха
синтакса	rdf:RDF	Коренски елемент RDF/ XML документа
синтакса	rdf:Description	Олакшава дефинисање „субјекта“ RDF тројке. У општој употреби је <code>rdf:Description about="URI to subject"&gt;</code> .
синтакса	rdf:ID	Овај атрибут пружа начин за скраћивање RDF синтаксе дозвољавањем употребе релативних URI идентификатора. Овај атрибут може да се користи заједно са атрибутом <code>xml:base</code> .
синтакса	rdf:about	Овај атрибут у суштини чува „субјекат“ тројке.
синтакса	rdf:parseType	Овај атрибут омогућава поједностављење синтаксе за креирање празних чворова.
синтакса	rdf:resource	
синтакса	rdf:nodeID	Овај атрибут подржава именовање празних чворова.

синтакса	rdf:datatype	Овај атрибут омогућава дефинисање типова података вредности објекта када се користе дословне вредности у RDF синтакси. Вредности типа података у складу су са типовима података за XMLS шему.
класа	rdf:Property	
класа	rdf:XMLLiteral	
својство	rdf:type	Овај елемент дозвољава експлицитно дефинисање типа ресурса који се описује и омогућава повезивање на основу више ресурса.
својство	rdf:value	Овај елемент омогућава дефинисање дословних вредности у контексту јединице мере.
својство	_n (n is any value > 1)	

RDF/XML изгледа и функционише као било који други XML документ. Пример за RDF/XML документ у Приказу 2.1 заснива се на метаподацима представљеним у Табели 2.2.

### Приказ 2.1

Основни RDF/XML фајл за *Аваншуре Хаклбери Фина*

```

1 <?xml version="1.0"?>
2 <rdf:RDF xmlns:rdf="http://www.
w3.org/1999/02/22-rdf-syntaxns#"
3 xmlns:dc="http://purl.
org/dc/elements/1.1/"
xmlns:bibo="http://purl.org/
ontology/bibo/">

```

```
4 <rdf:Description
rdf:about="http://ccn.loc.
gov/35020965">
5 <dc:title>The Adventures of
Huckleberry Finn</dc:title>
6 <dc:publisher
rdf:resource="http://id.loc.
gov/authorities/names/
n85242407"/>
7 <dc:subject
rdf:resource="http://id.loc.
gov/authorities/subjects/
sh2008110345"/>
8 <rdf:type
rdf:resource="bibo:Book"/>
9 </rdf:Description>
10 </rdf:RDF>
```

У Приказу 2.1 налази се једноставан RDF запис у коме се налазе четири тројке из Табеле 2.3. Хајде да прегледамо овај RDF/XML документ ред по ред. У првом реду видимо да је у питању један XML документ. У другом реду отвара се `rdf:RDF` елемент и то је основни елемент овог документа. Елемент `rdf:RDF` такође дефинише три простора имена, сопствени (`rdf`), простор имена за Dublin Core (`dc`) и простор имена за Библиографску онтологију (`bibo`). Ови простори имена користе се као предикати за тројке од петог до осмог реда. Елемент `rdf:Description` у четвртном реду успоставља субјекат синтакси тројки за редове од петог до осмог. У петом реду користи се дословна вредност као субјекат, док се у шестом, седмом и осмом реду користе референце ка ресурсима.

Овај једноставни RDF/XML документ може се погледати у форми графа уколико га укуцамо у виртуелни W3C валидатор. Измењени преглед овог графа приказан је у Шеми 2.5. Зарад боље прегледности, URL локације RDF/XML записа скраћене су до основних делова субјекта, предиката и објекта.

RDF Validator:  
[www.w3.org/RDF/Validator](http://www.w3.org/RDF/Validator)

Наша једноставна анализа овог RDF/XML формата за серијализацију не представља детаљан увод. Поред опширног и превише упрошћеног документа који смо навели, RDF/XML формат такође подржава и читав спектар стандарда који се заснивају на XML формату, као што су дефиниција и употреба ентитета, проширена употреба квалификованих имена (QNames)

за URI симплификацију, као и примена XML речника, као што су XML Datatype. Притом, постоје бројне пречице приликом креирања RDF докумената. Зарад разумљивости, примери представљени у овом поглављу прате опширнији модел. За више детаља о RDF синтакси консултујте RDF документацију на сајту W3C конзорцијума.

На крају ћемо истаћи и да RDF модел података садржи велики број елемената и атрибута који подржавају дефинисање група елемената који могу бити у одређеном редоследу (rdf:Seq) или без редоследа (rdf:Bag). Сваки од ових елемената садржи појединачне елементе у низу rdf:li елемената. RDF модел такође садржи збирке које представљају дефинисани број поља који може бити конфигуриран тако да нису дозвољене измене у саставу поља. То је значајно за креирање онтолошких структура у којима се за RDF модел дефинишу „неопходни и довољни“ елементи ресурса. Детаљније ћемо говорити о овоме када будемо анализирали Језик онтологије на вебу (OWL).

RDF модел подржава и креирање административних метаподатака користећи rdf:statement и сродне елементе. То се назива реификација и може да обухвати додавање екстерних речника како би се пружио одговарајући контекст. Више информација о томе налази се у Основама RDF модела.

## НАЧИНИ НОТАЦИЈЕ RDF МОДЕЛА – 3/N3, TURTLE, N-TRIPLES

N3 представља текстуалну серијализацију RDF модела, која није у XML формату и читљивија је људима, али истовремено прати и синтаксу која је лако читљива рачунарима. У Основама N3 стандарда пружен је опширан преглед начина на који се користи N3 код RDF модела, укључујући и дефинисање правила за RDF. Насупрот њему, тзв. Turtle<sup>27</sup> стандард је скраћена верзија N3 стандарда, која се фокусира на RDF исказе. N-Triples је формат који се фокусира на серијализацију тројки ред по ред. Из тог разлога, стандард N-Triple користи се за ширење RDF исказа, с обзиром на његову компактну синтаксу.

Основе N3 стандарда:

[www.w3.org/2000/10/swap/Primer.html](http://www.w3.org/2000/10/swap/Primer.html)

Познавање Turtle серијализације добар је пут ка упознавању RDF стандарда за упите SPARQL, зато што умногоме користе исту синтаксу.

---

27 Terse RDF Triple Language

## RDFa и JSON-LD: МИКРОФОРМАТИ И МИКРОПОДАЦИ

Док су стручњаци за метаподатке у библиотекама, архивима и музејима углавном усредсређени на изналагање свеобухватних решења за сложене проблеме са метаподацима, постоји и настојање да се системи метаподатака у самом документу, као што су Schema.org, RDFa, JSON-LD, Open Graph Protocol и многи други укључе у веб-податке библиотека, архива и музеја. Постоје бројне предности ових формата, укључујући једноставност, интероперабилност са популарним веб-сервисима попут Фејсбука, као и прилагодљивост.<sup>28</sup> На пример, OCLC<sup>29</sup> користи Schema.org метаподатке за WorldCat базу података, што пружа нове могућности девелоперима да креирају веб-сервисе богате подацима и омогућава претраживачима да користе библиографске податке за индексирање и проналажење података.<sup>30</sup>

Микроформате је лако искористити јер се заснивају на постојећим шемама за метаподатке (нпр. hCard, vCard) и начинима серијализације (нпр. HTML). Притом, ови стандарди су лако применљиви и с обзиром на то да се лако интегришу у веб-странице, представљају значајан први корак ка изједначавању разумевања веб-података између рачунара и људи.<sup>31</sup>

Међутим, са друге стране, микроподаци представљају оно што ће крајњи корисник видети, те зато немају подједнак утицај као суштинско реструктурирање метаподатака за „back-end“ системе. Како би микроподаци били употребљивији, потребно је да дође до ове кључне промене. Заправо, тренутно се микроподаци користе тако да одговарају појединачним потребама. На пример, Фејсбук Open Graph Protocol користи се свуда на вебу, а, између осталог, и за сајтове као што су LibraryThing и IMDB.

Schema.org:

<http://schema.org>

LibraryThing:

<http://librarything.org>

IMDB:

[www.imdb.com](http://www.imdb.com)

---

<sup>28</sup> Erik Mitchell, "Linked Data Publishing for Libraries, Archives, and Museums: What Is the Next Step?" *Journal of Web Librarianship*, forthcoming.

<sup>29</sup> Online Computer Library Center.

<sup>30</sup> Ted Fons, Jeff Penka and Richard Wallis, "OCLC's Linked Data Initiative: Using Schema.org to Make Library Data Relevant," *Information Standards Quarterly* 24, no. 2/3 (Spring/Summer 2012): 29–33, doi:10.3789/isqv24n2-3.2012.05.

<sup>31</sup> Liyang Yu, *A Developer's Guide to the Semantic Web* (Heidelberg; New York: Springer, 2011), 95.

Табела 2.5

Приказ графа једноставног записа, из Приказа 2.1

RDF тип елемента	RDF елемент	Општа сврха
класа	<code>rdfs:Resource</code>	коренска класа RDFS речника
класа	<code>rdfs:Class</code>	овај термин користи се за дефинисање класе у речнику
класа	<code>rdfs:Literal</code>	
класа	<code>rdfs:Datatype</code>	
својство	<code>rdfs:range</code>	
својство	<code>rdfs:domain</code>	
својство	<code>rdfs:subClassOf</code>	
својство	<code>rdfs:subPropertyOf</code>	
својство	<code>rdfs:label</code>	
својство	<code>rdfs:comment</code>	
повезано својство	<code>rdfs:seeAlso</code>	
повезано својство	<code>rdfs:isDefinedBy</code>	

Општа структура Turtle исказа прати синтаксу субјекат–предикат–објекат, али користећи структуру из једног реда. У Приказу 2.2 налази се RDF запис из Приказа 2.1, представљен преко Turtle синтаксе. Сваки исказ има свој ред и субјекат и предикат. Обратите пажњу на то да се свака синтакса завршава тачком и да се елемент `@prefix` користи за дефинисање Dublin Core (`dc`) простора имена у првом реду.

Приказ 2.2

RDF ресурс приказан у виду Turtle синтаксе

1 `@prefix dc: <http://purl.org/dc/elements/1.1/>.`

2 `@prefix madsrdf: <http://www.loc.gov/mads/rdf/v1#>.`

3 `<http://lcn.loc.gov/35020965>`  
`dc:title "The Adventures of`  
`Huckleberry Finn".`



4 <<http://lccn.loc.gov/35020965>>  
dc:publisher <<http://id.loc.gov/authorities/names/n85242407>>.

5 <<http://lccn.loc.gov/35020965>>  
dc:subject <<http://id.loc.gov/authorities/subjects/sh2008110345>>.

6 <<http://id.loc.gov/authorities/names/n85242407>> a  
madsrdf:Authority.

Turtle садржи и специјалне симболе, као што је слово „a“, који упућује на елемент `rdf:type` (<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>). Ова пречица омогућава креирање `rdf:type` синтаксе, као што се може видети у 6. реду Приказа 2.2. Постоје још неке пречице за Turtle, као што су тачка иarez која служи за повезивање синтакси у којима се само објекти мењају (<subject>;<predicate> <object>, <object>.) и празан чвор који се дефинише коришћењем угластих заграда [ ]. За више информација о исправној структури Turtle исказа погледајте W3C документацију или неки од препоручених приручника.

W3C: Turtle  
[www.w3.org/TR/turtle](http://www.w3.org/TR/turtle)

## RDFa

RDFa наводимо као вид серијализације RDF модела, зато што се у њему примењује скраћена верзија RDF модела која користи HTML атрибуте. На W3C сајту налази се детаљан приручник за RDFa који се обично користи за уграђивање метаподатака у HTML податке. Велики број структура код RDFa стандарда поклапа се са RDF моделом, као нпр. концепт типа, подршка за простор имена и структура за представљање предиката (нпр. `property`, `rel` и `rev`) и субјеката и објеката (нпр. атрибут *about*, у зависности од контекста).

Приручник за RDFa:  
[www.w3.org/TR/xhtml-rdfa-primer](http://www.w3.org/TR/xhtml-rdfa-primer)

У Приказу 2.3 налази се једноставан RDFa запис уграђен у HTML. У овом примеру искоришћен је елемент `span` за RDF исказе у контексту верзија са слободним текстом. У првом исказу, који идентификује наслов дела у трећем реду, атрибут `about` служи за то да усмери на субјекат, атрибут

rel служи као предикат, а дословна вредност у елементу *span* служи као објекат. У шестом реду постоји слична структура, али уместо коришћења дословне вредности после *span* елемента, користи се URI као објекат са атрибутом за садржај.

### Приказ 2.3

RDFa верзија HTML библиографског описа *Аваншџура Хаклбери Фина*

```
1 <div xmlns:dc="http://
  2  purl.org/dc/elements/1.1/"
  3  xmlns:bibo="http://purl.org/
  4  ontology/bibo/">
  5  2 <p>
  6  3 <span about="http://
  7    lccn.loc.gov/35020965"
  8    rel="dc:title">The Adventures
  9    of Huckleberry Finn</span>is a
 10  4 <span about="http://lccn.loc.
 11    gov/35020965" rel="bibo:Book"
 12    />book printed in 1884 by
 13  5 <span about="http://
 14    lccn.loc.gov/35020965"
 15    property="dc:publisher"
 16    content="http://id.loc.
 17    gov/authorities/names/
 18    n8524240"/>Chatto & Windus
 19    about
 20  6 <span about="http://
 21    lccn.loc.gov/35020965"
 22    property="dc:subject"
 23    content="http://id.loc.
 24    gov/authorities/subjects/
 25    sh2008110345"/>Huckleberry Finn
 26  7 </p>
 27  8 </div>
```

Користећи принципе веб-дизајна, било би једноставно претворити ове RDFa HTML документе у текст читљив како људима, тако и рачунарима. Иако постоји мноштво информација о RDFa, довољно је да овде наведемо само основе. Постоје бројни квалитетни приручници и упутства о овом

стандарду.<sup>32</sup> Алтернативни приступ коришћењу RDFa који се фокусира на уграђивање метаподатака на веб-страницу постиже се преко GRDDL<sup>33</sup> стандарда. Он се заснива на идеји да се контекстуални метаподаци, као што су информације о аутору и издавачу, и богати линкови најбоље раздвајају на детаљно структурисане XML фајлове који се могу модификовати за специфичне потребе путем XSLT трансформација.<sup>34</sup> Као ни RDFa, GRDDL стандард такође није у фокусу наше анализе, тако да нећемо улазити у детаљнија разматрања.

## JSON-LD

JSON-LD<sup>35</sup> документује се на интернету. Овај формат за серијализацију примењује одређена својства RDF апстрактног модела користећи JSON и представља један од бројних JSON – серијализованих RDF модела. JSON представља нотацију засновану на објекту која се користи на вебу. Популаран је због једноставне синтаксе и могућности да се користи као објекат без икаквих модификација. Међутим, недостатак овог формата је неусклађеност речника у паровима података кључ/вредност. На пример, у Приказу 2.4 налазе се подаци о књизи Марка Твена у обичном JSON формату. Иако је запис лако читљив, не постоје индикатори за речнике коришћене за креирање записа, а нису коришћени ни URI идентификатори који би олакшавали повезивање са спољашњим подацима.

JSON-LD документација:

<http://json-ld.org>

Списак RDF серијализација у JSON нотацији:

[www.w3.org/wiki/JSON%2BRDF](http://www.w3.org/wiki/JSON%2BRDF)

### Приказ 2.4

Основни JSON запис са библиографским подацима за књигу

```
1 {  
2 "The Adventures of Huckleberry  
Finn": {  
3 "is_a": "Printed Book",  
4 "is_written_by": "Twain, Mark",
```

---

32 Dathan, "HTML5 Microdata, Microformats, and RDFa Tutorials and Resources," WebsitesMadeRight.com (blog), May 1, 2011, <http://websitesmaderight.com/2011/05/html5-microdata-microformats-andrdfa-tutorials-and-resources>; Yu, A Developer's Guide to the Semantic Web.

33 Gleaning Resource Descriptions from Dialects of Languages.

34 Harry Halpin and Ian Davis, eds., "GRDDL Primer," W3C Working Group Note, June 28, 2007, [www.w3.org/TR/grddl-primer](http://www.w3.org/TR/grddl-primer).

35 JavaScript Object Notation for Linking Data.

```

5 "has_publication_date": "1884",
6 "has_length": "438 pages",
7 "is_published_by": "Chatto and
Windus",
8 "is_about": "Finn, Huckleberry
(Fictitious Character)
Fiction",
9 "is_about": "Runaway children,
Fiction"
10 }
11 }

```

Са друге стране, JSON-LD подржава коришћење URI идентификатора као референце за ресурс и поседује структуру за унос сличну као `rdf:type` у виду својства `@context`. Могућност рада са JSON-LD, који је у саставу програмског језика, драстично снижава препреке ка његовом усвајању јер нису потребне додатне библиотеке за рашчлањивање како би се приступило подацима. То је кључни фактор који утиче на усвајање отворених повезаних података.<sup>36</sup> JSON-LD представља радну верзију коју је развио W3C конзорцијум. Она се користи код апликационог програмског интерфејса (API) Дигиталне јавне библиотеке Америке. Више информација о томе можете наћи у документацији коју смо раније навели, а о овом стандарду говорићемо детаљније у трећем поглављу.

### Приказ 2.5

Једноставан JSON-LD запис са три дескриптивна поља за библиографски запис

```

1 {
2 "@context": "http://json-ld.
org/contexts/book.jsonld",
3 "@id": "http://lccn.loc.
gov/35020965",
4 "Title": {
5 "@id": "http://lccn.loc.
gov/35020965",

```

---

<sup>36</sup> Markus Lanthaler and Christian Gütl, "On Using JSON-LD to Create Evolvable RESTful Services," in WS-REST '12: Proceedings of the Third International Workshop on RESTful Design (New York: ACM, 2012), 25–32, doi:10.1145/2307819.2307827; Manu Sporny, Dave Longley, Gregg Kellogg, Markus Lanthaler and Niklas Lindström, "JSON-LD 1.0: A JSON-Based Serialization for Linked Data," W3C Last Call Working Draft, April 11, 2013, [www.w3.org/TR/json-ld-syntax](http://www.w3.org/TR/json-ld-syntax).

```

6 "name": "The Adventures of
Huckleberry Finn"
7 },
8 "is_a": {
9 "@id": "http://purl.org/
ontology/bibo#book",
10 "name": "Printed Book"
11 },
12 "is_written_by": {
13 "url": "http://id.loc.gov/
authorities/names/n79021164",
14 "name": "Twain, Mark"
15 }
16 }

```

## ПЕТА ОДЛИКА: РАЗМЕНА ОТВОРЕНИХ ПОВЕЗАНИХ ПОДАТАКА

За размену отворених повезаних података кључно је утврдити да је институција или заједница која их је објавила имала право на то. Објављивање отворених повезаних података на заједничкој платформи или преко агрегатора представља јединствен изазов јер захтева нове уговоре о пружању услуга и одређену техничку подршку и подршку заједнице, како би само подаци са отвореном лиценцом били доступни као повезани подаци.<sup>37</sup> Поред тога, националне и међународне организације, као што су Европеана и Дигитална јавна библиотека Америке, од кључног су значаја за стварање критичке масе за усвајање отворених повезаних података, тако да су ова питања и значајна и остварива. Кријејтив комонс лиценце за изражавање ауторских права (ccREL)<sup>38</sup> представљају заједнички отворени речник за изражавање права<sup>39</sup> који се користи у бројним пројектима за отворене повезане податке библиотека, музеја и архива.

Поред питања ауторских права, техничка страна размене отворених повезаних података укључује протоколе за упите попут SPARQL протокола и RDF језика за упите (SPARQL). SPARQL, који препоручује W3C конзорцијум, састоји се од скупа спецификација које одређују структуру упита и протокол за упите и примање података. SPARQL 1.1 подржава ре-

<sup>37</sup> Haslhofer et al., "Europeana RDF Store Report," 94.

<sup>38</sup> The Creative Commons Rights Expression Language.

<sup>39</sup> Hal Abelson, Ben Adida, Mike Linksvayer and Nathan Yergler, "ccREL: The Creative Commons Rights Expression Language," W3C Member Submission, May 1, 2008, [www.w3.org/Submission/ccREL](http://www.w3.org/Submission/ccREL).

зултате у JSON, CSV, TSV и XML формату. SPARQL се заснива на семантичким односима дефинисаним за отворене повезане податке и усмерен је на пружање „одговора“ уместо „докумената“. Као резултат тога, SPARQL омогућава детаљну претрагу графова у изворима отворених повезаних података и он сам проналази RDF податке, што значи да је SPARQL сам по себи нова структура отворених повезаних података.

SPARQL 1.1 преглед:

[www.w3.org/TR/sparql11-overview](http://www.w3.org/TR/sparql11-overview)

SPARQL је значајан елемент семантичког веба, а Европеана га користи за отворене повезане податке. За више информација о SPARQL упитима, препоручујем књигу *Повезани подаци* Хита и Бизера, као и Јову књигу о семантичком вебу.<sup>40</sup>

## ЗАКЉУЧАК

У овом поглављу бавили смо се основама отворених повезаних података и отворених повезаних речника и разматрали смо улогу пет основних одлика метаподатака у подржавању ових спецификација. Закључили смо да RDF модел, иако је незгоднији за људску употребу, има бројне предности за компјутерску обраду и као такав отвара могућности за нове начине стварања и ширења метаподатака и њихове употребе. Закључили смо и да речничке спецификације, као што су RDFS, SKOS и OWL омогућавају машинско читање отворених повезаних података и отворених повезаних речника.

Можда може деловати да није било потребе говорити о OAI-ORE, SPARQL и JSON-LD стандардима, али битно је разумети ове концепте за студије случаја у трећем поглављу. На пример, OAI је основа EDM модела, а JSON-LD је формат за серијализацију за DPLA API. Основне одлике о којима смо овде говорили имају паралелу у системима који су неопходни за њихову примену. RDF се уобичајено чува у бази података која се назива *triple store* (нпр. Apache Fuseki) која се састоји од исказа који описују ресурсе користећи екстерне вокабуларе, шеме метаподатака и онтологије. Претрага се врши помоћу SPARQL упита и серијализује се помоћу једног од неколико формата, укључујући RDF/XML, JSON-LD, или сажете верзије попут RDFa. Иако може деловати превише компликовано, кључна предност употребе отворених повезаних података је у томе што *triple store* база података може да чува податке о неограниченом броју врста ресурса или речника, елиминишући тиме проблеме ограничености у размени подата-

---

<sup>40</sup> Tom Heath and Christian Bizer, *Linked Data: Evolving the Web into a Global Data Space, Synthesis Lectures on the Semantic Web: Theory and Technology*, ed. James Hendler (San Rafael, California: Morgan & Claypool, 2011), doi:10.2200/S00334ED1V01Y-201102WBE001; Yu, *A Developer's Guide to the Semantic Web*.

ка који се уобичајено јављају у библиотечкој пракси. У трећем поглављу ослонићемо се на познавање концепата које смо представили у овом поглављу и основних одлика метаподатака како бисмо анализирали три система за отворене повезане податке.

*Превела с енглеској Таијана Домазећ*

## **Abstract**

In chapter 2 of Library Technology Reports (vol. 49, no. 5) "Library Linked Data: Research and Adoption" we explore the world of linked open data (LOD) and linked open vocabularies (LOV) through the lens of our five building blocks of metadata (data model, content rules, metadata schema, data serialization, and data exchange). This chapter provides a common foundation for understanding the technologies used in the case studies in chapter 3 and frames the issues and opportunities associated with LOD/LOV in LAM (libraries, archives, and museums) communities.