

Linked Data

STRUCTURED DATA ON THE WEB

DAVID WOOD
MARSHA ZAIDMAN
AND LUKE RUTH
WITH MICHAEL HAUSENBLAS



MANNING
Shelter Island

©Copyright 2014 by Dreamtech Press, 19-A, Ansari Road, Daryaganj, New Delhi-110002

Original English Language Edition ©Copyright by Manning Publications Co., USA.

This book may not be duplicated in any way without the express written consent of the publisher, except in the form of brief excerpts or quotations for the purpose of review. The information contained herein is for the personal use of the reader and may not be incorporated in any commercial programs, other books, databases, or any kind of software without written consent of the publisher. Making copies of this book or any portion thereof for any purpose other than your own is a violation of copyright laws.

This edition has been published by arrangement with the Manning Publications Co., USA. and published in India by Dreamtech Press, 19-A, Ansari Road, Daryaganj, New Delhi-110002.

Limits of Liability/disclaimer of Warranty : The author and publisher have used their best efforts in preparing this book. Manning Publications Co., USA and The author makes no representation or warranties with respect to the accuracy or completeness of the contents of this book, and specifically disclaim any implied warranties of merchantability or fitness for any particular purpose. There are no warranties which extend beyond the descriptions contained in this paragraph. No warranty may be created or extended by sales representatives or written sales materials. The accuracy and completeness of the information provided herein and the opinions stated herein are not guaranteed or warranted to produce any particular results, and the advice and strategies contained herein may not be suitable for every individual. Neither Manning Publications Co., USA, nor the author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Trademarks: All brand names and product names used in this book are trademarks, registered trademarks, or trade names of their respective holders. Manning Publications Co., USA is not associated with any product or vendor mentioned in this book.

ISBN: 978-93-5119-400-2

Edition : 2014

Printed at : Sanat Printers, Sonapat.

brief contents

PART 1	THE LINKED DATA WEB	1
	1 ■ Introducing Linked Data	3
	2 ■ RDF: the data model for Linked Data	27
	3 ■ Consuming Linked Data	60
PART 2	TAMING LINKED DATA	77
	4 ■ Creating Linked Data with FOAF	79
	5 ■ SPARQL—querying the Linked Data Web	99
PART 3	LINKED DATA IN THE WILD	123
	6 ■ Enhancing results from search engines	125
	7 ■ RDF database fundamentals	158
	8 ■ Datasets	178
PART 4	PULLING IT ALL TOGETHER	207
	9 ■ Callimachus: a Linked Data management system	209
	10 ■ Publishing Linked Data—a recap	233
	11 ■ The evolving Web	239

contents

foreword xiii
preface xv
acknowledgments xvii
about this book xix
about the cover illustration xxiii

PART 1 THE LINKED DATA WEB.....1

1 *Introducing Linked Data* 3

1.1 Linked Data defined 4

1.2 What Linked Data won't do for you 6

1.3 Linked Data in action 7

Freeing data 7 ▪ *Linked Data with Google rich snippets and Facebook likes* 8 ▪ *Linked Data to the rescue at the BBC* 9

1.4 The Linked Data principles 11

Principle 1: Use URIs as names for things 11 ▪ *Principle 2: Use HTTP URIs so people can look up those names* 12 ▪ *Principle 3: When someone looks up a URI, provide useful information* 12
Principle 4: Include links to other URIs 13

1.5 The Linking Open Data project 14

1.6 Describing data 15

- 1.7 RDF: a data model for Linked Data 18
- 1.8 Anatomy of a Linked Data application 20
 - Accessing a facility's Linked Data* 22
 - *Creating the user interface from Linked Data* 24
- 1.9 Summary 26

2 *RDF: the data model for Linked Data* 27

- 2.1 The Linked Data principles extend RDF 28
- 2.2 The RDF data model 33
 - Triples* 33
 - *Blank nodes* 35
 - *Classes* 36
 - *Typed literals* 37
- 2.3 RDF vocabularies 38
 - Commonly used vocabularies* 39
 - *Making your own vocabularies* 42
- 2.4 RDF formats for Linked Data 43
 - Turtle—human-readable RDF* 44
 - *RDF/XML—RDF for enterprises* 46
 - *RDFa—RDF in HTML* 49
 - *JSON-LD—RDF for JavaScript Developers* 52
- 2.5 Issues related to web servers and published Linked Data 54
- 2.6 File types and web servers 56
 - When you can configure Apache* 57
- 2.7 When you have limited control over Apache 57
- 2.8 Linked Data platforms 58
- 2.9 Summary 58

3 *Consuming Linked Data* 60

- 3.1 Thinking like the Web 61
- 3.2 How to consume Linked Data 62
- 3.3 Tools for finding distributed Linked Data 64
 - Sindice* 64
 - *SameAs.org* 65
 - *Data Hub* 65
- 3.4 Aggregating Linked Data 66
 - Aggregating some Linked Data from known datasets* 66
 - *Getting Linked Data and RDF from web pages using browser plug-ins* 70
- 3.5 Crawling the Linked Data Web and aggregating data 73
 - Using Python to crawl the Linked Data Web* 73
 - *Creating HTML output from your aggregated RDF* 75
- 3.6 Summary 76

PART 2 TAMING LINKED DATA.....77

4 *Creating Linked Data with FOAF* 79

- 4.1 Creating a personal FOAF profile 80
 - Introducing the FOAF vocabulary* 81 ▪ *Method I: manual creation of a basic FOAF profile* 82 ▪ *Enhancing a basic FOAF profile* 83
 - Method II: automated generation of a FOAF profile* 85
- 4.2 Adding more content to a FOAF profile 88
- 4.3 Publishing your FOAF profile 90
- 4.4 Visualization of a FOAF profile 91
- 4.5 Application: linking RDF documents using a custom vocabulary 93
 - Creating a wish list vocabulary* 93 ▪ *Creating, publishing, and linking the wish list document* 94 ▪ *Adding wish list items to our wish list document* 95 ▪ *Explanation of our bookmarklet tool* 97
- 4.6 Summary 98

5 *SPARQL—querying the Linked Data Web* 99

- 5.1 An overview of a typical SPARQL query 100
- 5.2 Querying flat RDF files with SPARQL 101
 - Querying a single RDF data file* 102 ▪ *Querying multiple RDF files* 104 ▪ *Querying an RDF file on the Web* 106
- 5.3 Querying SPARQL endpoints 107
- 5.4 Types of SPARQL queries 109
 - The SELECT query* 109 ▪ *The ASK query* 111 ▪ *The DESCRIBE query* 111 ▪ *The CONSTRUCT query* 112
 - SPARQL 1.1 Update* 113
- 5.5 SPARQL result formats (XML, JSON) 113
- 5.6 Creating web pages from SPARQL queries 115
 - Creating the SPARQL query* 116 ▪ *Creating the HTML page* 117 ▪ *Creating the JavaScript for the table* 118
 - Creating JavaScript for the map* 119
- 5.7 Summary 122

PART 3 LINKED DATA IN THE WILD123

6 *Enhancing results from search engines* 125

- 6.1 Enhancing HTML by embedding RDFa 126
 - RDFa markup using FOAF vocabulary* 129 ▪ *Using the HTML span attribute with RDFa* 132 ▪ *Extracting Linked Data from a FOAF-enhanced HTML document* 133
- 6.2 Embedding RDFa using the GoodRelations vocabulary 134
 - An overview of the GoodRelations vocabulary* 134 ▪ *Enhancing HTML with RDFa using GoodRelations* 137 ▪ *A closer look at selections of RDFa GoodRelations* 143 ▪ *Extracting Linked Data from GoodRelations-enhanced HTML document* 145
- 6.3 Embedding RDFa using the schema.org vocabulary 148
 - An overview of schema.org* 148 ▪ *Enhancing HTML with RDFa Lite using schema.org* 150 ▪ *A closer look at selections of RDFa Lite using schema.org* 152 ▪ *Extracting Linked Data from a schema.org enhanced HTML document* 154
- 6.4 How do you choose between using schema.org or GoodRelations? 155
- 6.5 Extracting RDFa from HTML and applying SPARQL 155
- 6.6 Summary 156

7 *RDF database fundamentals* 158

- 7.1 Classifying RDF databases 158
 - Selecting an RDF database systems* 160 ▪ *RDF databases versus RDBMS* 161 ▪ *Benefits of RDF database systems* 166
- 7.2 Transforming spreadsheet data to RDF 167
 - A basic RDF conversion of MS Excel* 167 ▪ *Transforming MS Excel to Linked Data* 169 ▪ *Finding RDF converter tools* 171
- 7.3 Application: collecting Linked Data in an RDF database 171
 - Outlining the process* 171 ▪ *Using Python to aggregate our data sources* 172 ▪ *Understanding the output* 175
- 7.4 Summary 177

8 *Datasets* 178

- 8.1 Description of a Project 179
 - Creating a DOAP profile* 180 ▪ *Using the DOAP vocabulary* 182

- 8.2 Documenting your datasets using VoID 186
 - The Vocabulary of Interlinked Datasets* 186
 - *Preparing a VoID file* 187
- 8.3 Sitemaps 190
 - Non-semantic sitemaps* 190
 - *Semantic sitemaps* 192
 - Enabling discovery of your site* 194
- 8.4 Linking to other people's data 195
- 8.5 Examples of using owl:sameAs to interlink datasets 200
- 8.6 Joining Data Hub 203
- 8.7 Requesting outgoing links from DBpedia to your dataset 204
- 8.8 Summary 206

PART 4 PULLING IT ALL TOGETHER207

9 *Callimachus: a Linked Data management system* 209

- 9.1 Getting started with Callimachus 211
- 9.2 Creating web pages using RDF classes 212
 - Adding data to Callimachus* 212
 - *Telling Callimachus about your OWL class* 213
 - *Associating a Callimachus view template to your class* 214
- 9.3 Creating and editing class instances 216
 - Creating a new note* 218
 - *Creating a view template for a note* 219
 - *Creating an edit template for notes* 220
- 9.4 Application: creating a web page from multiple data sources 222
 - Making and querying Linked Data from NOAA and EPA* 222
 - Creating a web page to contain the application* 224
 - *Creating JavaScript to retrieve and display Linked Data* 226
 - *Bringing it all together* 229
- 9.5 Summary 232

10 *Publishing Linked Data—a recap* 233

- 10.1 Preparing your data 234
- 10.2 Minting URIs 235
- 10.3 Selecting vocabularies 236
- 10.4 Customizing vocabulary 237
- 10.5 Interlinking your data to other datasets 237

- 10.6 Publishing your data 238
- 10.7 Summary 238

11 *The evolving Web* 239

- 11.1 The relationship between Linked Data and the Semantic Web 239
 - Demonstrated successes* 243
- 11.2 What's coming 245
 - Google extended rich snippets* 245 ▪ *Digital accountability and transparency legislation* 245 ▪ *Impact of advertising* 246
 - Enhanced searches* 246 ▪ *Participation by the big guys* 246
- 11.3 Conclusion 247

- appendix A Development environments* 249
- appendix B SPARQL results formats* 252
 - glossary* 258
 - index* 265