

---

# Contents

<b>I</b>	<b>Introduction</b>	<b>1</b>
<b>1</b>	<b>Linked Data &amp; the Semantic Web Standards</b>	<b>3</b>
	<i>Aidan Hogan</i>	
1.1	Introduction . . . . .	3
1.2	Semantic Web . . . . .	6
1.3	Resource Description Framework (RDF) . . . . .	10
1.3.1	RDF Terms . . . . .	10
1.3.2	RDF Triples and Graphs . . . . .	12
1.3.3	RDF Vocabulary . . . . .	16
1.3.4	RDF Syntaxes . . . . .	18
1.4	RDF Semantics, Schemata and Ontologies . . . . .	20
1.4.1	RDF Semantics . . . . .	20
1.4.2	RDF Schema (RDFS) . . . . .	23
1.4.3	Web Ontology Language (OWL) . . . . .	25
1.5	Querying RDF with SPARQL . . . . .	31
1.5.1	Query Types . . . . .	33
1.5.2	Dataset Clause and Named Graphs . . . . .	34
1.5.3	Query Clause . . . . .	35
1.5.4	Solution Modifiers . . . . .	39
1.5.5	Towards SPARQL 1.1 . . . . .	40
1.6	Linked Data . . . . .	41
1.6.1	The Early Semantic Web on the Web . . . . .	42
1.6.2	Linked Data Principles and Best Practices . . . . .	43
1.6.3	Linking Open Data . . . . .	44
<b>2</b>	<b>Aligning Ontologies of Linked Data</b>	<b>49</b>
	<i>Rahul Parundekar, Craig A. Knoblock, and José Luis Ambite</i>	
2.1	Introduction . . . . .	49
2.2	Linked Data Sources with Heterogeneous Ontologies . . . . .	51
2.3	Finding Alignments Across Ontologies . . . . .	51
2.3.1	Source Preprocessing . . . . .	51
2.3.2	Aligning Atomic Restriction Classes . . . . .	52
2.3.3	Aligning Conjunctive Restriction Classes . . . . .	53
2.3.4	Eliminating Implied Alignments . . . . .	55
2.3.5	Finding Concept Coverings . . . . .	57

2.3.6	Curating Linked Data . . . . .	59
2.4	Results . . . . .	60
2.4.1	Representative Examples of Atomic Alignments . . . . .	60
2.4.2	Representative Examples of Conjunctive Alignments . . . . .	61
2.4.3	Representative Examples of Concept Coverings . . . . .	62
2.4.4	Outliers . . . . .	62
2.4.5	Precision and Recall of Country Alignments . . . . .	63
2.4.6	Representative Alignments from Other Domains . . . . .	63
2.5	Related Work . . . . .	67
2.6	Conclusion . . . . .	68
<b>3</b>	<b>Architecture of Linked Data Applications</b> . . . . .	<b>69</b>
	<i>Benjamin Heitmann, Richard Cyganiak, Conor Hayes, and Stefan Decker</i>	
3.1	Introduction . . . . .	69
3.2	An Empirical Survey of RDF-based Applications . . . . .	71
3.2.1	Research Method of the Survey . . . . .	71
3.2.2	Findings . . . . .	73
3.2.3	Discussion of Threats to Validity . . . . .	76
3.3	Empirically-grounded Conceptual Architecture . . . . .	77
3.3.1	Architectural Style of the Conceptual Architecture . . . . .	77
3.3.2	Decomposing the Surveyed Applications into Components . . . . .	78
3.3.3	Description of Components . . . . .	79
3.4	Implementation Challenges . . . . .	84
3.4.1	Integrating Noisy and Heterogeneous Data . . . . .	84
3.4.2	Mismatch of Data Models between Components . . . . .	84
3.4.3	Distribution of Application Logic across Components . . . . .	85
3.4.4	Missing Guidelines and Patterns . . . . .	85
3.5	Analysis of an Example Application . . . . .	86
3.6	Future Approaches for Simplifying Linked Data Application Development . . . . .	88
3.6.1	More Guidelines, Best Practices, and Design Patterns . . . . .	88
3.6.2	More Software Libraries . . . . .	88
3.6.3	Software Factories . . . . .	89
3.7	Related Work . . . . .	89
3.7.1	Empirical Surveys . . . . .	90
3.7.2	Architectures for Semantic Web Applications . . . . .	90
3.8	Conclusions . . . . .	91
<b>II</b>	<b>Centralized Query Processing</b> . . . . .	<b>93</b>

<b>4</b>	<b>Mapping Relational Databases to Linked Data</b>	<b>95</b>
	<i>Juan F. Sequeda and Daniel P. Miranker</i>	
4.1	Introduction . . . . .	95
4.2	Scenarios . . . . .	96
4.2.1	Direct Mapping and ETL . . . . .	96
4.2.2	Manual R2RML Mapping and ETL . . . . .	97
4.2.3	Direct Mapping as Ontology with Ontology Alignment . . . . .	98
4.2.4	Automatic Mapping with Wrapper . . . . .	99
4.3	W3C RDB2RDF Standards . . . . .	100
4.3.1	Running Example . . . . .	100
4.3.2	Direct Mapping . . . . .	100
4.3.3	R2RML: RDB to RDF Mapping Language . . . . .	103
4.4	Semantic Integration Architecture . . . . .	106
4.4.1	$\mathcal{DM}_{ont}$ : Direct Mapping as Ontology . . . . .	108
4.4.2	Wrapper Systems . . . . .	111
4.4.2.1	Information Preservation . . . . .	111
4.4.2.2	Query Preservation . . . . .	111
4.4.2.3	Ultrawrap . . . . .	112
4.5	Conclusions . . . . .	114
<b>5</b>	<b>Efficient Query Processing in RDF Databases</b>	<b>117</b>
	<i>Andrey Gubichev and Thomas Neumann</i>	
5.1	Introduction . . . . .	117
5.2	Storing and Indexing RDF Data . . . . .	118
5.2.1	Relational-based Approach . . . . .	119
5.2.2	Graph-based Approach . . . . .	123
5.3	Query Processing and Query Optimization of SPARQL Queries . . . . .	124
5.3.1	Translating SPARQL Queries . . . . .	124
5.3.2	Join Ordering . . . . .	126
5.3.3	Cardinality Estimation . . . . .	130
5.3.4	Physical Operators: Joins, Path Traversals . . . . .	134
5.4	Open Issues and Conclusion . . . . .	135
<b>6</b>	<b>Evaluating SPARQL Queries over Linked Data Streams</b>	<b>137</b>
	<i>Jean-Paul Calbimonte and Oscar Corcho</i>	
6.1	Data Stream and Event Query Processing . . . . .	139
6.1.1	Data Stream and Event Models . . . . .	139
6.1.2	Continuous Query Processing and Windows . . . . .	142
6.1.3	Streaming Data & Event Processing Engines . . . . .	143
6.2	SPARQL Extensions for Querying RDF Streams . . . . .	145
6.2.1	RDF Stream Model . . . . .	145
6.2.2	SPARQL Extensions for Event and Data Streams . . . . .	146
6.2.3	SPARQL Extensions Semantics . . . . .	149
6.3	RDF Stream Query Processing . . . . .	151
6.3.1	Query Processing Strategies . . . . .	151

6.3.2	Mappings from Streaming Data to RDF . . . . .	153
6.3.3	Query Rewriting over Diverse Streaming Data Technologies . . . . .	155
6.4	Discussion and Lessons Learned . . . . .	158
6.4.1	Lessons Learned . . . . .	160
6.4.2	Future Directions . . . . .	160
<b>III</b>	<b>Parallel Query Processing</b>	<b>163</b>
<b>7</b>	<b>SPARQL Query Processing in the Cloud</b>	<b>165</b>
	<i>Francesca Bugiotti, Jesús Camacho-Rodríguez, François Goasdoué, Zoi Kaoudi, Ioana Manolescu, and Stamatis Zampetakis</i>	
7.1	Introduction . . . . .	165
7.2	Classification of RDF Data Management Platforms in the Cloud . . . . .	167
7.2.1	Cloud-based RDF Storage . . . . .	167
7.2.2	Cloud-based SPARQL Query Processing . . . . .	168
7.2.3	RDF Reasoning in the Cloud . . . . .	169
7.3	AMADA: RDF Data Repositories in the Amazon Cloud . . . . .	170
7.3.1	Amazon Web Services . . . . .	170
7.3.2	General Architecture . . . . .	173
7.4	RDF Storing and SPARQL Querying within AMADA . . . . .	175
7.4.1	Data Model and Running Example . . . . .	175
7.4.2	Answering Queries from the Index . . . . .	177
7.4.3	Selective Indexing Strategies . . . . .	179
7.5	Implementation Details . . . . .	184
7.6	Experimental Evaluation . . . . .	185
7.6.1	Experimental Setup and Datasets . . . . .	185
7.6.2	Indexing Time and Costs . . . . .	186
7.6.3	Querying Time and Costs . . . . .	189
7.6.4	Scalability . . . . .	190
7.6.5	Experiments Conclusion . . . . .	191
7.7	Summary and Perspectives . . . . .	191
<b>8</b>	<b>The Bigdata<sup>®</sup> RDF Graph Database</b>	<b>193</b>
	<i>Bryan Thompson, Mike Personick, and Martyn Cutcher</i>	
8.1	Introduction . . . . .	193
8.2	Bigdata Database Architecture . . . . .	194
8.2.1	Deployment Models . . . . .	194
8.2.2	Scale-Up Architecture . . . . .	199
8.2.3	Scale-Out Architecture . . . . .	205
8.3	RDF Database Architecture . . . . .	213
8.3.1	Resource Description Framework . . . . .	213
8.3.2	Database Schema for RDF . . . . .	214
8.3.3	SPARQL Query Processing . . . . .	216

8.3.4	Reification Done Right . . . . .	224
8.4	Related Work . . . . .	227
8.4.1	Clustered Graph Databases . . . . .	227
8.4.2	Key-Value and Map/Reduce Systems . . . . .	230
8.4.3	Graph Mining Platforms . . . . .	234
8.4.4	Analysis . . . . .	235
8.5	Conclusion . . . . .	236
<b>9</b>	<b>Experiences with Virtuoso Cluster RDF Column Store</b>	<b>239</b>
	<i>Peter Boncz, Orri Erling, and Minh-Duc Pham</i>	
9.1	Virtuoso Cluster . . . . .	239
9.1.1	Architecture of a Server Process . . . . .	240
9.1.2	Differences with Virtuoso 6 . . . . .	242
9.2	Query Evaluation . . . . .	244
9.2.1	Vectoring . . . . .	245
9.2.2	Cluster Tuning Parameters . . . . .	245
9.2.3	Micro Benchmarks . . . . .	246
9.3	Experimental Evaluation . . . . .	250
9.3.1	BSBM Results to Date . . . . .	250
9.3.2	Cluster Configuration . . . . .	251
9.3.3	Bulk Loading RDF . . . . .	251
9.3.4	Notes on the BI Workload . . . . .	253
9.3.5	BSBM Benchmark Results . . . . .	254
9.4	Conclusion . . . . .	258
<b>IV</b>	<b>Distributed Query Processing</b>	<b>261</b>
<b>10</b>	<b>Linked Data Query Processing Based on Link Traversal</b>	<b>263</b>
	<i>Olaf Hartig</i>	
10.1	Introduction . . . . .	263
10.2	Theoretical Foundations . . . . .	265
10.2.1	Data Model . . . . .	266
10.2.2	Full-Web Semantics for SPARQL . . . . .	267
10.3	Principles and Characteristics of LTBQE . . . . .	268
10.3.1	Live Exploration . . . . .	269
10.3.2	Integrated Execution . . . . .	271
10.3.3	Combining Live Exploration and Integrated Execution	271
10.4	An Iterator Based Implementation . . . . .	272
10.4.1	Introduction to the Implementation Approach . . . . .	273
10.4.2	Characteristics of the Implementation Approach . . . . .	276
10.4.3	Selecting Query Plans for the Approach . . . . .	278
10.4.4	Improving Performance of the Approach . . . . .	281
10.4.5	Alternative Implementation Approaches . . . . .	282
10.5	Summary . . . . .	282

<b>11 Semantic Navigation on the Web of Data</b>	<b>285</b>
<i>Valeria Fionda, Claudio Gutierrez, and Giuseppe Pirrò</i>	
11.1 Introduction . . . . .	285
11.2 The NAUTLOD Navigational Language . . . . .	287
11.2.1 Preliminaries . . . . .	288
11.2.2 Syntax of NAUTLOD . . . . .	289
11.2.3 Semantics of NAUTLOD . . . . .	290
11.2.4 Evaluation Cost and Complexity of NAUTLOD . . . . .	293
11.3 NAUTLOD: Algorithms and Implementation . . . . .	294
11.3.1 Evaluating NAUTLOD Expressions . . . . .	294
11.3.2 <code>swget</code> Scripts . . . . .	296
11.3.3 <code>swget</code> Architecture . . . . .	297
11.4 Some Application Scenarios . . . . .	299
11.4.1 The <code>swget</code> API and Command Line . . . . .	299
11.4.2 The <code>swget</code> GUI . . . . .	300
11.4.3 The <code>swget</code> Web Portal . . . . .	302
11.5 Related Research . . . . .	303
11.6 Concluding Remarks . . . . .	306
<b>12 Index-Based Source Selection and Optimization</b>	<b>311</b>
<i>Jürgen Umbrich, Marcel Karnstedt, Axel Polleres, and Kai-Uwe Sattler</i>	
12.1 Introduction . . . . .	311
12.2 Source Selection Approaches . . . . .	313
12.2.1 Generic Query Processing Model & Assumption for Source Selection . . . . .	313
12.2.2 Source-Selection Approaches . . . . .	314
12.3 Data Summaries . . . . .	316
12.3.1 Multidimensional Histograms . . . . .	317
12.3.2 QTree . . . . .	319
12.3.3 Construction and Maintenance . . . . .	322
12.4 Importance of Hashing . . . . .	323
12.4.1 Hash Functions . . . . .	324
12.4.2 Comparison of Hash Functions . . . . .	326
12.5 Source Selection . . . . .	327
12.5.1 Triple Pattern Source Selection . . . . .	327
12.5.2 Join Source Selection . . . . .	328
12.5.3 Result Cardinality Estimation and Source Ranking . . . . .	332
12.6 Practicability . . . . .	334
12.7 Conclusion . . . . .	336
<b>13 P2P-Based Query Processing over Linked Data</b>	<b>339</b>
<i>Marcel Karnstedt, Kai-Uwe Sattler, and Manfred Hauswirth</i>	
13.1 Introduction . . . . .	339
13.2 General Architecture . . . . .	342

13.3	A Distributed Index . . . . .	343
13.3.1	DHTs for Index Management . . . . .	343
13.3.2	Indexing RDF Data . . . . .	346
13.4	Query Processing . . . . .	348
13.4.1	Logical Algebra & Operators . . . . .	348
13.4.2	Physical Operators & Query Execution . . . . .	350
13.4.3	Query Execution . . . . .	356
13.4.4	Query Planning . . . . .	360
13.5	Other Existing Approaches . . . . .	363
13.6	Summary . . . . .	367
<b>14</b>	<b>Federated Query Processing over Linked Data</b>	<b>369</b>
	<i>Peter Haase, Katja Hose, Ralf Schenkel, Michael Schmidt, and Andreas Schwarte</i>	
14.1	Introduction . . . . .	369
14.2	State of the Art . . . . .	371
14.3	Optimization Techniques . . . . .	375
14.3.1	Federated Query Processing Model . . . . .	376
14.3.2	Source Selection . . . . .	377
14.3.3	Join Ordering . . . . .	378
14.3.4	Exclusive Groups . . . . .	379
14.3.5	Bind Joins . . . . .	379
14.3.6	Alternative Join Techniques . . . . .	381
14.4	Benchmarking Federation Techniques . . . . .	381
14.4.1	The FedBench Benchmark Suite . . . . .	383
14.4.2	SPLODGE . . . . .	385
14.4.3	Additional Work on Benchmarking . . . . .	386
14.5	Summary and Outlook . . . . .	387
<b>V</b>	<b>Reasoning over RDF Data</b>	<b>389</b>
<b>15</b>	<b>On the Use of Abstract Models for RDF/S Provenance</b>	<b>391</b>
	<i>Irini Fundulaki, Giorgos Flouris, and Vassilis Papakonstantinou</i>	
15.1	Introduction . . . . .	391
15.2	Representation Models for Provenance . . . . .	392
15.2.1	Concrete Representation Models . . . . .	393
15.2.2	Abstract Representation Models . . . . .	395
15.2.3	Handling Changes in Abstract Models . . . . .	397
15.2.4	Alternative Abstract Models . . . . .	397
15.3	Abstract Provenance Models: Formalities . . . . .	398
15.3.1	Basic Concepts . . . . .	398
15.3.2	Defining Abstract Provenance Models . . . . .	399
15.4	Algorithms . . . . .	401
15.4.1	Annotation . . . . .	401
15.4.2	Evaluation . . . . .	402

15.4.3	Adding Quadruples . . . . .	403
15.4.4	Deleting Quadruples . . . . .	403
15.4.5	Changing the Semantics . . . . .	405
15.5	Implementation of an Abstract Provenance Model . . . . .	405
15.5.1	Basic Schema . . . . .	405
15.5.2	Partitioned Schema . . . . .	406
15.6	Experimental Evaluation . . . . .	407
15.6.1	Experimental Setting . . . . .	407
15.6.2	Experimental Results . . . . .	408
15.7	Conclusions . . . . .	411
<b>16</b>	<b>Incremental Reasoning on RDF Streams</b>	<b>413</b>
	<i>Daniele Dell’Aglia and Emanuele Della Valle</i>	
16.1	Introduction . . . . .	413
16.2	Basic Concepts . . . . .	414
16.2.1	RDF Stream . . . . .	414
16.2.2	RDFS+ . . . . .	415
16.3	Running Example . . . . .	415
16.4	Incremental Materialization Algorithm for RDF Streams . . . . .	419
16.4.1	Assumptions and Preliminary Definitions . . . . .	419
16.4.2	Maintenance Program Generation . . . . .	422
16.4.3	Execution of the Maintenance Program in the IMaRS Window . . . . .	427
16.5	Related Works . . . . .	428
16.6	Evaluation . . . . .	432
16.7	Conclusions and Future Works . . . . .	434
<b>VI</b>	<b>Linked Data Interfaces</b>	<b>437</b>
<b>17</b>	<b>Linked Data Services</b>	<b>439</b>
	<i>Sebastian Speiser, Martin Junghans, and Armin Haller</i>	
17.1	Introduction . . . . .	439
17.2	Scenario . . . . .	440
17.3	Information Services . . . . .	441
17.4	LIInked Data Services (LIDS) . . . . .	443
17.5	Describing Linked Data Services . . . . .	445
17.5.1	Relation to Source Descriptions in Information Integration Systems . . . . .	446
17.5.2	Describing LIDS Using RDF and SPARQL Graph Patterns . . . . .	447
17.6	Interlinking Data with LIDS . . . . .	448
17.7	Related Work . . . . .	450
17.8	Implementation and Evaluation . . . . .	451
17.8.1	Realization of Scenario . . . . .	451
17.8.2	Implementing and Interlinking Linked Data Services . . . . .	454

17.8.2.1	Implementing LIDS Services . . . . .	455
17.8.2.2	Interlinking Existing Data Sets with LIDS . . . . .	456
17.9	Conclusion . . . . .	456
<b>18</b>	<b>Using read-write Linked Data for Application Integration</b>	<b>459</b>
	<i>Arnaud J Le Hors and Steve Speicher</i>	
18.1	Introduction . . . . .	459
18.2	The Integration Challenge . . . . .	460
18.3	What Would Success Look Like? . . . . .	461
18.4	The Solution . . . . .	462
18.5	Linked Data as an Application Integration Framework . . . . .	463
18.6	The Linked Data Platform (LDP) . . . . .	463
18.6.1	Introduction to the Linked Data Platform (LDP) . . . . .	465
18.6.2	Linked Data Platform Resources (LDPR) . . . . .	465
18.6.3	Linked Data Platform Container (LDPC) . . . . .	469
18.6.4	Retrieving the Non-member Properties of an LDPC . . . . .	471
18.6.5	Creating and Adding a Resource to an LDPC . . . . .	471
18.6.6	Inlined Member Resources . . . . .	472
18.6.7	Updating an LDPR . . . . .	473
18.6.8	Deleting an LDPR . . . . .	473
18.6.9	Creating an LDPC . . . . .	473
18.6.10	Updating an LDPC . . . . .	474
18.6.11	Deleting an LDPC . . . . .	474
18.6.12	Adding a Binary Resource to an LDPC . . . . .	475
18.6.13	Membership Subject and Predicate of an LDPC . . . . .	477
18.6.14	Paging . . . . .	478
18.6.15	Ordering . . . . .	480
18.7	Observations . . . . .	481
18.8	LDP Limitations . . . . .	481
18.9	Conclusion . . . . .	482
	<b>Bibliography</b>	<b>485</b>
	<b>Index</b>	<b>537</b>