Table of Contents

Preface xi

Part I. The Basics

1. Mind the Semantic Gap 3
What Is Semantic Data Modeling? 4
Why Develop and Use a Semantic Data Model? 7
Bad Semantic Modeling 8
Avoiding Pitfalls 10
Breaking Dilemmas 11

2. Semantic Modeling Elements 13
General Elements 14
Entities 14
Relations 16
Classes and Individuals 17
Attributes 19
Complex Axioms, Constraints, and Rules 21
Terms 22
Common and Standardized Elements 23
Lexicalization and Synonymy 24
Instantiation 25
Meaning Inclusion and Class/Relation Subsumption 26
Part-Whole Relation 26
Semantic Relatedness 28
Mapping and Interlinking Relations 28

Documentation Elements 29
Summary 32

3. Semantic and Linguistic Phenomena..... 35
Ambiguity 35
Uncertainty 37
Vagueness 38
Rigidity, Identity, Unity, and Dependence 41
Symmetry, Inversion, and Transitivity 42
Closed- and Open-World Assumptions 43
Semantic Change 43
Summary 46

4. Semantic Model Quality..... 49
Semantic Accuracy 50
Completeness 52
Consistency 54
Conciseness 55
Timeliness 57
Relevancy 57
Understandability 58
Trustworthiness 59
Availability, Versatility, and Performance 60
Summary 61

5. Semantic Model Development..... 63
Development Activities 63
Setting the Stage 64
Deciding What to Build 67
Building It 68
Ensuring It's Good 69
Making It Useful 69
Making It Last 70
Vocabularies, Patterns, and Exemplary Models 71
Upper Ontologies 71
Design Patterns 72
Standard and Reference Models 74
Public Models and Datasets 74
Semantic Model Mining 76
Mining Tasks 76
Mining Methods and Techniques 79

Summary 84

Part II. The Pitfalls

6. Bad Descriptions..... 87
Giving Bad Names 87
Setting a Bad Example 89
Why We Give Bad Names 90
Pushing for Clarity 91
Omitting Definitions or Giving Bad Ones 92
When You Need Definitions 93
Why We Omit Definitions 94
Good and Bad Definitions 95
How to Get Definitions 96
Ignoring Vagueness 97
Vagueness Is a Feature, Not a Bug 99
Detecting and Describing Vagueness 100
Not Documenting Biases and Assumptions 107
Keeping Your Enemies Close 107
Summary 108

7. Bad Semantics..... 109
Bad Identity 109
Bad Synonymy 110
Bad Mapping and Interlinking 113
Bad Subclasses 115
Instantiation as Subclassing 115
Parts as Subclasses 117
Rigid Classes as Subclasses of Nonrigid Classes 118
Common Superclasses with Incompatible Identity Criteria 119
Bad Axioms and Rules 119
Defining Hierarchical Relations as Transitive 119
Defining Vague Relations as Transitive 121
Complementary Vague Classes 121
Mistaking Inference Rules for Constraints 122
Summary 123

8. Bad Model Specification and Knowledge Acquisition..... 125
Building the Wrong Thing 125
Why We Get Bad Specifications 126

How to Get the Right Specifications 128
Bad Knowledge Acquisition 133
Wrong Knowledge Sources 134
Wrong Acquisition Methods and Tools 140
A Specification and Knowledge Acquisition Story 146
Model Specification and Design 146
Model Population 149
Summary 153

9. Bad Quality Management 155
Not Treating Quality as a Set of Trade-Offs 155
Semantic Accuracy Versus Completeness 156
Conciseness Versus Completeness 156
Conciseness Versus Understandability 157
Relevancy to Context A Versus Relevancy to Context B 157
Not Linking Quality to Risks and Benefits 158
Not Using the Right Metrics 160
Using Metrics with Misleading Interpretations 160
Using Metrics with Little Comparative Value 162
Using Metrics with Arbitrary Value Thresholds 162
Using Metrics That Are Actually Quality Signals 164
Measuring Accuracy of Vague Assertions in a Crisp Way 165
Equating Model Quality with Information Extraction Quality 166
Summary 167

10. Bad Application 169
Bad Entity Resolution 169
How Entity Resolution Systems Use Semantic Models 170
When Knowledge Can Hurt You 171
How to Select Disambiguation-Useful Knowledge 172
Two Entity Resolution Stories 178
Bad Semantic Relatedness 181
Why Semantic Relatedness Is Tricky 182
How to Get the Semantic Relatedness You Really Need 183
A Semantic Relatedness Story 184
Summary 187

11. Bad Strategy and Organization 189
Bad Strategy 189
What Is a Semantic Model Strategy About? 190
Buying into Myths and Half-Truths 192

Underestimating Complexity and Cost 193
Not Knowing or Applying Your Context 195
Bad Organization 196
Not Building the Right Team 196
Underestimating the Need for Governance 200
Summary 202

Part III. The Dilemmas

12. Representation Dilemmas..... 205
Class or Individual? 205
To Subclass or Not to Subclass? 208
Attribute or Relation? 211
To Fuzzify or Not to Fuzzify? 212
What Fuzzification Involves 212
When to Fuzzify 220
Two Fuzzification Stories 222
Summary 225

13. Expressiveness and Content Dilemmas..... 227
What Lexicalizations to Have? 227
How Granular to Be? 231
How General to Be? 233
How Negative to Be? 235
How Many Truths to Handle? 237
How Interlinked to Be? 238
Summary 241

14. Evolution and Governance Dilemmas..... 243
Model Evolution 243
Remember or Forget? 244
Run or Pace? 244
React or Prevent? 246
Knowing and Acting on Your Semantic Drift 249
Model Governance 255
Democracy, Oligarchy, or Dictatorship? 255
A Centralization Story 257
Summary 258

15. Looking Ahead..... 261
The Map Is Not the Territory 261
Being an Optimist, but Not Naïve 262
Avoiding Tunnel Vision 263
Avoiding Distracting Debates 263
Semantic Versus Nonsemantic Frameworks 263
Symbolic Knowledge Representation Versus Machine Learning 265
Doing No Harm 266
Bridging the Semantic Gap 267

Bibliography..... 269

Glossary..... 291

Index..... 295