

Table of Contents

Introduction	1
What Does This Book Cover?	1
Who Is This Book For?	2
What You Need to Use This Book	2
How Is This Book Structured?	2
Conventions	5
Support, Errata, and P2P	6
Tell Us What You Think	7
Chapter 1: Getting Started with XML Schemas	9
Why do we Need Schemas?	10
Background to XML Schemas	10
The W3C XML Schema Recommendation	12
Getting Started with XML Schemas	13
The Difference Between Simple and Complex Types	14
Simple Types	14
Complex Types	15
Element Declarations	16
Global versus Local Element Declarations	17
Element Occurrence Indicators	20
Value Constraints on Element Content – Default and Fixed Content	22
Attribute Declarations	23
Global versus Local Attribute Declarations	24
Occurrence of Attributes	24
Value Constraints on Attributes	25
Annotations	27
Validating an Instance Document	29
How the XML Schema Recommendation Specifies Validity	30
Validating with XSV	31
Validating a Schema	32
Validating an Instance Document	34
Example Schema: Delivery Receipt	36
Summary	40

Chapter 2: Datatype Basics	43
Datatypes in XML – An Overview	44
Properties of XML Schema Datatypes	45
Value Spaces	45
Lexical Spaces	46
Facets	46
Primitive versus Derived Datatypes	47
Built-in Types versus User-Defined	48
Kinds of Datatype	48
Atomic Datatypes	48
List Datatypes	49
Union Datatypes	50
Namespaces for Built-in Datatypes	52
Built-in Primitive Types	53
String Types	54
string	54
anyURI	54
NOTATION	55
QName	56
Encoded Binary Types	56
boolean	56
hexBinary	57
base64Binary	57
Numeric Types	57
decimal	57
float	58
double	58
Date/Time Types	59
duration	59
dateTime	60
date	61
time	62
gYearMonth	62
gYear	62
gMonthDay	63
gMonth	63
gDay	63
Built-in Derived Types	64
string Types	66
normalizedString	66
token	66
language	66
name, NCName	67
ID	67
IDREF	68
IDREFS	68
NMTOKEN	68
NMTOKENS	68
ENTITY	69
ENTITIES	69
decimal Types	69
integer	69
nonPositiveInteger, nonNegativeInteger	69
positiveInteger, negativeInteger	70
long, int, short, byte	70
unsignedLong, unsignedInt, unsignedShort, unsignedByte	70
Summary	71

Chapter 3: Creating Content Models	75
Complex Type Definitions	76
Compositors	78
Combining Compositors	80
Repeating Compositors	80
Named Model Groups	81
Repeating Model Groups	83
Attribute Groups	84
Element Content Specifications	85
Element Content	86
Mixed Content	88
Empty	89
Any Content	89
Nil Values	92
Example Schema: Contacts	93
Schema Review	99
Features We Still Have to Meet	99
Identity Constraints	100
Namespace Support	101
Deriving Complex Types	101
Using Constructs from Other Schemas	101
The 13 Schema Constructs	102
Summary	103
Chapter 4: Deriving New Types	105
Facets of Datatypes	105
Fundamental Facets	106
Equality	106
Order	107
Bounds	108
Cardinality	108
Numeric/Non-Numeric	109
Constraining Facets	109
length, minLength, maxLength	109
whiteSpace	111
pattern	112
enumeration	113
minExclusive, minInclusive, maxExclusive, maxInclusive	113
totalDigits, fractionDigits	114
Constraining Facets of Built-in Datatypes	115
Primitive Datatypes	116
Derived Datatypes	117
Regular Expressions	118
XML Schema Regexes	119
Basic Grammar	119
Atoms	121
Meta Characters	121
Character Classes	122
Character Class Expressions	122
Character Class Escapes	124
Quantifiers	130

Table of Contents

Regex Syntax Comparison	131
Unicode	132
Perl	133
Regular Expressions Summary	134
The xsi:type Attribute	135
Datatype Inheritance	135
Type Hierarchy	136
Deriving Simple Types	137
Deriving Complex Types	138
Restrictions Upon Element Content	138
Deriving by Extension	139
Derived by Restriction	142
Substitution Groups	143
Abstract Types	144
Controlling Datatype Derivation	145
The fixed Attribute	145
The final Attribute	145
The finalDefault Attribute	146
The block Attribute	146
The blockDefault Attribute	146
Summary	146
Chapter 5: Some Useful Datatypes	149
Sample Custom Datatypes	150
Some Simple Basic Types	150
Countries	150
Country Codes (ISO 3166)	151
Currency Codes (ISO 4217)	153
Language Codes (ISO 639)	155
Example – Describe a Country	156
Internet Datatypes	157
IP Addresses	158
Domain Names	159
Uniform Resource Identifiers	160
People	163
Names	163
Gender	164
Telephone Numbers	165
Postal Addresses	168
Street Addresses	169
Regions	170
Postal Codes	172
The postalAddress Datatype – Revised	173
Example – NAFTA and British Addresses	175
Geodetic Locations	178
Summary	180

Chapter 6: XML Schemas and Namespaces	183
XML Namespace Refresher	185
Why We Need XML Namespaces	185
How We Declare and Use Namespaces in XML Documents	185
Declaring a Namespace	186
The Scope of a Namespace	187
Namespace Defaulting	189
No Namespace	191
Uniqueness of Attributes	192
The Three Namespaces of XML Schema	192
The XML Schema Namespace	193
The XML Schema Datatypes Namespace	193
The XML Schema Instance Namespace	194
Defining Namespace for Markup	195
Creating Vocabularies that Do Not Belong to a Namespace	195
Creating Vocabularies that Belong to a Namespace	196
Looking Behind Namespace Qualification	198
Understanding Infoset Contributions to Namespace Qualification	199
Requiring All Elements and Attributes Be Qualified	201
Summary of Rules to Remember when Using Namespaces	206
Local Control of Namespace Qualification	207
Models and Issues for Local versus Global Declarations	208
Russian Doll Design	209
Salami Slice Design	210
Characteristics of Russian Doll Design	212
Characteristics of Salami Slice Design	213
Venetian Blind Model	214
Going Deeper into the Venetian Blind Model	216
Issues with Complex Types and the Venetian Blind Model	219
We Can Only Associate Each Element Name with One Type	219
Complex Types and Model Groups as Building Blocks	221
Summary	225
Chapter 7: Schema Design Fundamentals	227
Prescriptive versus Descriptive Models	228
Processing XML Documents	228
Naming Conventions	229
Example of Two Elements with the Same Name and Different Content Models	232
Consistency, Order, and Occurrence	233
Consistency, Order, and Occurrence in Document Structures	234
Effects of Consistency, Order, and Occurrence Upon Processing Requirements	236
Consistency, Order, and Occurrence in Schemas	236
Re-Use	237
Declaring Order and Occurrence	241
Requiring Elements in a Mixed Model	244

Table of Contents

Grouping and Hierarchy	245
Grouping Related Items of Information in Documents	246
Grouping Same Information Items in Documents	247
Advantages of Grouping for Processing	248
Creating Hierarchical Structures in Schemas	249
Choice and Options	249
A Choice of Fixed Values	250
Handling Choices for Simple Content Using Enumeration	250
Choice of Structures	251
Handling Choices of Content Models	252
Choice Groups Versus Substitution Groups	253
Choices Between Elements and Their Content Models	254
Flexible Combinations of Order and Occurrence	258
A Note About Attributes for Distinguishing Types of Element	263
Summary	264
 Chapter 8: Creating Schemas From Multiple Documents	 267
Why Create Modular Schemas	268
Modularization of XHTML	269
How We Can Re-Use Information Across Schemas	270
Schemas from the Same Namespace or No Namespace	273
Schemas from a Different Target Namespace	280
How Many Namespaces Should We Use	284
Heterogeneous Namespace Design	285
Homogeneous Namespace Design	288
Chameleon Design	290
Which Approach Is Best?	291
When We Should Hide Or Expose Namespaces	293
Should Our Default Namespace be the targetNamespace or XML Schema Namespace	294
Name Collisions	297
Altering Schema Constructs	299
Methods of Altering Markup that We Import and Include	300
Deriving New Types and Element Content Models	300
Redefine	303
Substitution Groups	306
Forcing Substitution with Abstract Elements and Types	308
Allowing for Extension of Our Model	309
Substitution Groups	309
Using Unknown Content	311
Protecting Our Markup – Controlling Re-Use	313
Enforcing Namespace Qualification	313
Preventing Derivation of Types	314
Modularizing Schemas	315
Creating Our Own Modularized Schemas	315
Summary	318
 Chapter 9: Identity Constraints, Normalization, and Document Fragments	 321
Identity Constraints	322
Locating Nodes with XPath	325
Unique Values	329
Keys and Key References	332

Maintaining Stable, Secure, and Accurate Data	335
Clarifying Database Terms	337
Data Redundancy	337
Functional Dependencies	338
Properties of Fields	338
Atomicity	339
Data Integrity	339
Entity Integrity	340
Domain Integrity	341
Referential Integrity	342
User-Defined Integrity	343
Normal Forms	343
First Normal Form	344
Second Normal Form	345
Third Normal Form	346
Domain-Key Normal Form	347
Applying Normalization to XML	347
First Grade XML Normalization	347
Second Grade XML Normalization	348
Third Grade XML Normalization	348
XML Domain/key Norm (XML DKNF)	348
Working with Document Fragments	349
Working with Fragments	349
Problems with References and Uniqueness	350
Problems with Multiple Schemas	352
Context	353
Summary	353
Chapter 10: Schema and XSLT	357
Creating Documentation	358
Schema Information	360
Summarizing Components	361
Retrieving the Target Namespace	362
Summarizing List Attributes	363
Navigating Schemas	364
Navigating Between Components	365
Navigating Between schemas	369
Creating a Form	375
Accessing Schemas from Instances	375
Generating Forms	376
Creating Labels	380
Creating Form Fields	381
Navigating Model Groups	384
Summary	391

Chapter 11: XML System Modeling	393
Principles for Data Modeling	394
A Simplified Modeling Approach	394
Objects versus Instances	396
Magazine Publisher Goes XML	397
Establishing System Requirements	397
Magazine Publisher Background	399
Expect the Unexpected	399
Identify Key Players	399
Requirements Definition and Analysis	400
UML and XML Modeling	400
Business Process Analysis	402
Identifying Business Domain Objects	404
Use Case Diagram	405
System Decomposition and Abstraction	407
Information Modeling	408
Cardinality	409
Kinds of Relationship	409
Primary Object List	410
Future Proofing and Scope of the Model	414
Secondary Object List	414
Sequence Diagrams	416
Defining Properties	419
Distinguishing Business Rules from Application Rules	420
Custom Datatypes	423
Business Rules	423
Process Model	425
Iterative Development	426
Data Model	426
Namespaces, Schemas, and Instance Documents	426
Creating Schema and Instance Documents	427
Representing Relationships in XML Schema	433
Composition	433
Aggregation	433
Inheritance	435
Association	435
Global Declarations	435
Guidelines for Schema Design	436
Testing the Data Model	438
Summary	438
Chapter 12: Creating XML Schema for an Existing Database	441
The Business Problem	441
Technologies to Be Used	441
The Relational Database Schema	442

Migrating our Database to XML	443
Determining the Document's Scope	443
Representational or Transactional	443
Determining the Content	443
Representing Type Information	444
Creating a Traversal Path	444
Why Is A Traversal Path Important?	444
Traversing Our First Example	445
Creating the Schema	445
Creating User-Defined Datatypes	446
Lookup Tables	446
Common User-Defined Type Definitions	447
Modeling the Logical Schema Entities	448
Modeling the Logical Schema Attributes	448
Elements or Attributes	448
Modeling the Logical Schema Relationships	449
Using Parent-Child Relationships	449
Using key-keyref(s) Relationships	449
Orphaned Elements	450
A More Complex Example	452
Document Scope	452
The Traversal Path	453
The Logical Schema Entities	453
The Logical Schema Attributes	454
The Logical Schema Relationships	454
Orphans	455
Summary	457
 Chapter 13: W3C XML Schemas for Document Management	 459
Introduction	459
Document Schemas versus Data Schemas	460
What XML Documents Model and What They Leave Out	461
Case Study: Physics Press Goes Online	461
Managing Document Schema Development	462
The Schema Design Process: An Overview	463
Document Analysis	463
Analyzing Structures	463
Identifying Datatypes	466
Schema Design: Strategic Issues	467
Choosing an Appropriate Level of Granularity	467
Coping with Legacy Data	469
How Prescriptive Should Our Schema Be?	470
Smallest Licensable Unit	472
Many Specific or Few General Elements	474
Extensibility and Future Proofing	475
Generic Elements	476
Documentation and Usage Guidelines	477
Schema Design: Tactical Issues	477
Using Standard Modules	477
Data Types	478
Element Naming	479
Section Schemes	480

Table of Contents

Managing the Schema Design Phase	482
Turning the Design Into a Schema	483
Testing the Schema	490
Documentation	491
Managing Ongoing Development	492
Summary	492
 Chapter 14: Schematron and Other Schema Technologies	 493
 Schematron	 493
A Sample Schematron Schema	495
Compiling a Basic Validator	496
Compiling and Running an XML Validator	497
Compiling and Running an HTML Validator	498
More Schematron	499
Name	499
Diagnostics	500
Phases	500
Formatting/Documentation	503
Abstract/Extends	503
Key	505
Complementary Features	506
XML Schema is Grammar-based – Schematron is Rule-based	506
Schematron Allows Arbitrary Logical Constraints	506
Specify Root Elements	507
"Expansion Slots" Add User Specified Documentation and Processing Elements	507
Schematron Runs in any Environment that Supports XSLT	507
Schematron supports XSLT/XPath Constraints of any Complexity	508
Co-location	508
Why Co-locate?	509
Illustration	509
Code Notes	512
Limitations of Using Schematron with XML Schema	514
Names versus Types	514
Namespace Re-mapping	516
 Schematron and XSLT 2.0	 516
Sample Using MSXML 4 Extensions	516
Topologi – An Interactive Schematron Development Tool	520
 Competitive Schema Technologies	 523
RELAX NG	523
RELAX NG Versus XML Schemas and DTDs	524
Why RELAX NG?	527
Examplotron	528
More Examplotron	530
Why use Examplotron?	531
ASN.1	531
Who will Use it?	533
 Summary	 534

Chapter 15: E-Commerce Case Study	537
Introduction	537
Automating E-Commerce	538
Project Scenario	539
SOAP – An Overview	539
SOAP Message Structure	539
SOAP Envelope	540
SOAP Header	540
SOAP Body	541
SOAP Fault	542
Some SOAP Links	542
Project Requirements	543
Main Use Case	543
Design Goals	545
Architecture	545
Communication with the Customer	545
Communication Between the Bank and Partner	545
Design	546
Conversation Sequence Diagram	547
Transactions and Messages	548
Data Structures	549
Some Tables from the Database	549
The Message Schema	550
Sample Messages	555
LoanRequest	555
LoanRequestResponse	556
LoanResult	556
LoanResultResponse	557
Using Schematron with XML Schema	557
Non-XML Schema Constraints	557
Looking Ahead – A Web Services View	560
WSDL	560
UDDI	560
Some Web Services Links	561
Summary	561
Chapter 16: Schema-Based Programming	563
Declarative Programming	564
Formal Tools for Program Development	564
Implementing Applications with SBP	565
Model – View – Controller	565
Model	566
View	567
Controller	567

Table of Contents

SBP Schema	567
Model	568
View	569
Controller	570
Implementing an SBP Application	572
Scheme of Processing	572
Initial Download Actions	573
User Interface Creation	573
Controller Activation	574
Event Processing	578
Implementing the Model with Petri Nets	579
Petri Net Markup Language	581
General Structure of a PNML Document	581
Places	582
Transitions	583
Arcs	583
XSLT Stylesheet for PNML	583
Identifying States that Gain and Lose Tokens	583
Controlling the Transformation	585
Rich Tokens	587
Sample Application	589
Model	591
View	593
Controller	596
Define an Event	596
Write the Initialize Function	596
Writing the Event Handler	597
Future Directions	599
Summary	601
Appendix A: Schema Element and Attribute Reference	603
all	603
annotation	604
any	605
anyAttribute	606
appinfo	607
attribute	607
attributeGroup	608
choice	609
complexContent	610
complexType	610
documentation	612
element	612
extension	614
field	615
group	615
import	616
include	617
key	617
keyref	618
list	618
notation	619
redefine	619
restriction	620
schema	622
selector	623
sequence	623
simpleContent	624
simpleType	624
union	625
unique	626

XML Schema Instance Attributes	626
Appendix B: Schema Datatypes Reference	629
XML Schema Built-in Datatypes	630
Constraining Facets	636
length	636
minLength	637
maxLength	637
pattern	638
enumeration	639
whiteSpace	639
maxInclusive	640
maxExclusive	641
minExclusive	641
minInclusive	642
totalDigits	642
fractionDigits	643
Appendix C: UML Reference	649
Classes and Objects	649
Relationships	650
Dependency	650
Association	650
Aggregation	651
Composition	651
Multiplicity	651
Naming an Association	652
Inheritance	652
Multiple Inheritance	652
States	652
Object Interactions	653
Components	654
Appendix D: XML Schema Tools	657
Parsers & XML Development Kits (XDKs)	657
Xerces-c, Xerces-j (Apache)	659
XML4C, XML4J (IBM)	660
MSXML4 and XDK (Microsoft)	660
XML Schema Processor and XDKs (Oracle)	661
Schema Checkers	662
XML Schema Quality Checker (IBM)	662
XML Schema Validator	662
Authoring Tools	663
Turbo XML (TIBCO Extensibility)	663
XML Spy 3.5 and 4.0 (Altova)	664
Schema Conversion Tools	666

Appendix E: Bibliography and Further Reading	669
URLs	669
Data Modeling/UML	669
Date/Time Resources	669
Regular expressions	670
SBP Resources	670
Standards	670
Tools	672
Unicode	673
Web Services	673
Wrox Resources	673
XML	673
XML Schema Alternatives	673
Books	674

