

Preface

“Hey, you got peanut butter on my chocolate bar!”

“No way, you got chocolate in my peanut butter!”

If you consider data management (DM) as peanut butter and XML as chocolate—and you like the combination—then you will easily understand what we are attempting to explain in this book. If you don’t like the combination of peanut butter and chocolate, then pick your own favorite flavor combination. We like the yin and yang analogy, but you should go for some combination that works together. These analogies have helped the literally thousands who we have assisted or have seen us present on XML and DM topics over the past few years.

You may be confused by XML hype but nevertheless you are intrigued as to what it—XML, that is—is all about. You understand the basics of what XML can do, but you can’t seem to find any guide as to how best to make use of it. Well look no further! You won’t be coding much XML yourself, but as a director/strategic thinker/manager, you must understand your organizational requirements and conceive of how best to use XML to support the organization. You need to gain an understanding quickly of how to use XML to architect solutions to specific data-engineering challenges. The material we present is based on dozens of person-years of effort, creating and implementing the advanced DM concepts that we describe.

Your technical background also permits you to understand the importance of good DM practices to the organization. It has been a constant source of frustration that you have been unable to improve support of the organizational DM function; many have not managed to put together an effective approach given the current economic climate. The question is also, What should be done right now? What DM tools and techniques can you apply effectively in today's environment? Our extensive experience in these and related areas allows us to describe the new XML-enabled DM direction.

One question remains—how to decide what to read among the vast array of titles on the subject. There are (for example) at least four other titles that might be competing for your attention:

- *XML for Data Architects*
- *XML: A Manager's Guide*
- *XML Data Management*
- *Professional XML Meta Data*

These texts are either very technical or more narrowly focused on specific subjects. This book was necessary to describe how XML will make your job easier, give you increased DM capabilities, or save your organization resources.

Well, as a result of the happy mixing of XML and data management, by reading this book you will discover that significant synergies exist between XML and DM. XML and DM go together like yin and yang! Between the two of us authors, we have more than a decade of XML experience and—working on various Data Blueprint (<http://datablueprint.com>) projects—more than a half century of combined DM experience. XML, while not a silver bullet, does offer significant support to DM functions—support that will dramatically alter the level of investment required to produce positive return on investment. In short, XML will significantly lower DM implementation costs in a way that has not yet been imagined except by a very few organizations. This book will permit yours to be one of them.

Thank you for reading. Please enjoy your learning experience and do get back to us with any suggestions or corrections.

Peter Aiken and David Allen
Richmond, VA, April 2004
peter@datablueprint.com
mda@datablueprint.com

Contents

Preface	v
Chapter 1 XML and DM Basics	1
Introduction	1
The DM Challenge	2
Definitions	4
Data and Information	4
Metadata	5
Traditional Views of Metadata	7
DM Overview	8
Data Program Coordination	8
Enterprise Data Integration	10
Data Stewardship	11
Data Development	12
Data Support Operations	13
Data Asset Use	13
Investing in Metadata/Data Management	13
Typical Systems Evolution	14
XML Integration	16
Data Integration/Exchange Challenges	17
Managing Joan Smith's Metadata	18

XML Hype: Management by Magazine	20
Two Examples of XML in Context	23
Internet Congestion and Application Efficiency	23
Information Location	24
XML & DM Interaction Overview	26
Management of Unstructured Data	27
Expanded DM Roles	29
Preparation of Organizational Data for E-Business	31
What XML Is Not: XML Drawbacks and Limitations	31
Chapter Summary	32
References	32

Chapter 2 **XML from the Builder's Perspective: Using XML to Support DM** **35**

Chapter Overview	35
XML Builder's Overview	36
XML Terms	36
XML Parser/XML Processor	38
HTML Drawbacks	40
XML "Rules"	41
XML Usage in Support of DM: Builder's Perspective	48
Integration at the Browser	48
Integration via Hub and Spoke	50
B2B Example	53
Legacy Application Conversion	55
XML Conversion	57
Metadata Management Example	62
Chapter Summary	68
References	69

Chapter 3 **XML Component Architecture (as it relates to DM)** **71**

Introduction	71
XML Design Considerations	73
XML Design Goals	73
What XML Should Not Be Used For	81
XML Component Architecture (Parts & Pieces)	86
The XML Community	86
XML Component Organization	86
XML Namespaces	89
DTD: Document Type Definition	90
XML Schema	94

DOM: Document Object Model	96
XPath	98
XLink: XML Linking	99
XSL and XSLT	104
RDF: Resource Description Framework	110
SOAP: Simple Object Access Protocol	112
WSDL: Web Services Definition Language	114
UDDI: Universal Description, Discovery, and Integration	115
ADML: Architecture Description Markup Language	119
Conclusion	120

Chapter 4 **XML and Data Engineering** **123**

Introduction	123
Typical XML First Steps	125
Engineering XML-Based Data Structures as Part of an Organizational Data Architecture	126
If Software Companies Made Bridges	126
Metadata Engineering Analyses	134
XML and Data Quality	137
XML and Metadata Modeling	137
Data Structure Problem Difficulties	139
Engineering Roles	141
Measuring Data Engineering Tasks	143
XML, Security, and Data Engineering	145
SAML—Security Assertions Markup Language	146
XML Signatures	147
XKMS—The XML Key Management Services/System	147
Data Mapping Case Study	148
Project Planning Metadata	152
Extracting Metadata from Vendor Packages	154
Chapter Summary	159
References	160

Chapter 5 **Making and Using XML: The Data Managers' Perspective** **161**

Introduction	161
Input	162
XML Editors	162
CASE Technologies	164
Extracting Metadata From Legacy Systems	168
Processing XML	173

X Contents

XML Integration Servers	173
XML Mediation Servers	174
XML Repository Servers	176
Outputting XML	178
XML Converters	178
Generating XML Automatically or Semiautomatically	180
Data Layers/Data Services	182
Data Management Maturity Measurement (DM3)	183
Chapter Summary	185
References	186

Chapter 6 **XML Frameworks** **187**

Introduction	187
Framework Advantages	188
Shared Vocabulary—A Standardized Data Format	189
Standardize Processes	190
Connect as Many Organizations as Possible to Increase Value	191
Logical Hub-and-Spoke Model: Standards and Processes	192
Standardized Methods for Security and Scalability	193
Frameworks Lower Barrier to Entry	194
Commonly Available Transactions	194
RosettaNET	196
ebXML	200
Microsoft Offerings: BizTalk and .NET	205
BizTalk	205
.NET	209
Industry-Specific Initiatives	211
Acord	211
Envera	213
Common Themes and Services	217
Conclusion	221

Chapter 7 **XML-Based Portal Technologies and Data-Management Strategies** **223**

Chapter Overview	223
Portal Hype	224
The Need: Legacy Code Maintenance Burden	225
Aiding Implementation of Information-Engineering Principles with XML-Based Portal Architectures	228
Clarifying Excitement Surrounding XML-Based Portals (XBPs)	235
XML-Based Portal Technology	241

XML-Based Architectural Enhancements	247
Better Architectural Flexibility	247
Better Architectural Evolvability/Maintenance	248
Enhanced Integration Opportunities	250
Standards-Based Integration	250
More Integration Depth	251
Wider Integration Scope	251
More Rapid Implementation	252
Extending Data-Management Technologies/ Data-Management Product Examples	253
Selected Product Examples	256
Newly Important and Novel Data-Preparation Opportunities	259
Understanding Legacy Structures	260
XBPs and Data-Quality Engineering	261
Creating a Transitional Data Model	264
Greater Business and System-Reengineering Opportunities: Reduction of Maintenance Burden Strategies	265
Get Rid of Expensive-to-Maintain Code	266
Increased Integration Creates Demand for Portal Services Instead of Coded Applications	267
Conclusion	268
References	269

Chapter 8 **Focusing XML and DM on Enterprise Application
Integration (EAI)** **271**

Introduction	271
What Is It About XML That Supports EAI?	276
Flexibility and Structure	277
XML Messaging	277
Domain-Specific Languages	279
The Pundits Speak	280
EAI Basics	281
What Is Integration?	281
EAI Components	282
EAI Motivation	284
EAI Past and Current Focus	288
Generalized Approach to Integration	290
Resource Efficiencies	291
Engineering-Based Approach	291
Key Factor: Scalability = EAI Success	293
EAI Challenges	295
Lesson One: All About Data Integration	297

Lesson Two: Start Small 298
 Lesson Three: Core Technologies Lack EAI Support 299

Conclusion 300
References 301

Chapter 9 **XML, DM, and Reengineering 303**

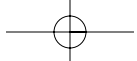
Introduction 303
Two Types of Reengineering 304
 The Broad Definition of Reengineering 304
 Business Process Reengineering 307
 The Relationship Between SR and BPR 310
How XML + DM Facilitates Reengineering Efforts 311
Chapter Summary 314
References 315

Chapter 10 **Networks of Networks, Metadata, and the Future 317**

Introduction 317
A Different Understanding of Data and Its Metadata 318
The Internet Metaphor 323
Internal Organizational Structure 325
 Internal Organizational Data Interchange and the Internet Model 325
 The Use of XML-Based Metadata 326
Industry Structure 328
 XML and Intra-Industry Communication 330
Inter-Industry Structure 332
 Challenges Related to Connecting Industries 334
Bringing It Together: Observations About the Internet Metaphor 336
Conclusion 337

Chapter 11 **Expanded Data-Management Scope 339**

Introduction 339
Thought Versus Action 342
Understanding Important Data Structures as XML 345
 Example: Capacity Management with XML 347
 Example: Legacy Application Maintenance Reduction 349
 Example: Business Engineering 350
Resolving Differing Priorities 352
Producing Innovative XML-Based IT Savings 354
Increasing Scope and Volume of Data Management 357
Greater Payoff for Preparation 358



**Understanding the Growth Patterns in Your Operational
Environment 358**
Chapter Summary 359
References 359

Glossary of Acronyms 361

Index 365

