Contents

Preface 15

Chapter 1: Data Management 17

1. Introduction 17
2. Essential Concepts 18
   - 2.1 Data 18
   - 2.2 Data and Information 20
   - 2.3 Data as an Organizational Asset 20
   - 2.4 Data Management Principles 21
   - 2.5 Data Management Challenges 23
   - 2.6 Data Management Strategy 31
3. Data Management Frameworks 33
   - 3.1 Strategic Alignment Model 33
   - 3.2 The Amsterdam Information Model 34
   - 3.3 The DAMA-DMBOK Framework 35
   - 3.4 DMBOK Pyramid (Aiken) 39
   - 3.5 DAMA Data Management Framework Evolved 40
4. DAMA and the DMBOK 43
5. Works Cited / Recommended 46

Chapter 2: Data Handling Ethics 49

1. Introduction 49
2. Business Drivers 51
3. Essential Concepts 52
   - 3.1 Ethical Principles for Data 52
   - 3.2 Principles Behind Data Privacy Law 53
   - 3.3 Online Data in an Ethical Context 56
   - 3.4 Risks of Unethical Data Handling Practices 56
   - 3.5 Establishing an Ethical Data Culture 60
   - 3.6 Data Ethics and Governance 64
4. Works Cited / Recommended 65

Chapter 3: Data Governance 67

1. Introduction 67
   - 1.1 Business Drivers 70
   - 1.2 Goals and Principles 71
   - 1.3 Essential Concepts 72
2. Activities 79
   - 2.1 Define Data Governance for the Organization 79
   - 2.2 Perform Readiness Assessment 79
   - 2.3 Perform Discovery and Business Alignment 80
   - 2.4 Develop Organizational Touch Points 81
   - 2.5 Develop Data Governance Strategy 82
   - 2.6 Define the DG Operating Framework 82
   - 2.7 Develop Goals, Principles, and Policies 83
   - 2.8 Underwrite Data Management Projects 84
   - 2.9 Engage Change Management 85

2.10 Engage in Issue Management 86
2.11 Assess Regulatory Compliance Requirements 87
2.12 Implement Data Governance 88
2.13 Sponsor Data Standards and Procedures 88
2.14 Develop a Business Glossary 90
2.15 Coordinate with Architecture Groups 90
2.16 Sponsor Data Asset Valuation 91
2.17 Embed Data Governance 91
3. Tools and Techniques 92
3.1 Online Presence / Websites 92
3.2 Business Glossary 92
3.3 Workflow Tools 93
3.4 Document Management Tools 93
3.5 Data Governance Scorecards 93
4. Implementation Guidelines 93
4.1 Organization and Culture 93
4.2 Adjustment and Communication 94
5. Metrics 94
6. Works Cited / Recommended 95

Chapter 4: Data Architecture 97
1. Introduction 97
1.1 Business Drivers 99
1.2 Data Architecture Outcomes and Practices 99
1.3 Essential Concepts 101
2. Activities 108
2.1 Establish Data Architecture Practice 109
2.2 Integrate with Enterprise Architecture 113
3. Tools 114
3.1 Data Modeling Tools 114
3.2 Asset Management Software 114
3.3 Graphical Design Applications 114
4. Techniques 114
4.1 Lifecycle Projections 114
4.2 Diagramming Clarity 115
5. Implementation Guidelines 115
5.1 Readiness Assessment / Risk Assessment 116
5.2 Organization and Cultural Change 117
6. Data Architecture Governance 118
6.1 Metrics 118
7. Works Cited / Recommended 119

Chapter 5: Data Modeling and Design 121
1. Introduction 121
1.1 Business Drivers 123
1.2 Goals and Principles 123
1.3 Essential Concepts 124
2. Activities 149
2.1 Plan for Data Modeling 149

2.2 Build the Data Model 150
2.3 Review the Data Models 155
2.4 Maintain the Data Models 155
3. Tools 155
3.1 Data Modeling Tools 155
3.2 Lineage Tools 156
3.3 Data Profiling Tools 156
3.4 Metadata Repositories 156
3.5 Data Model Patterns 156
3.6 Industry Data Models 157
4. Best Practices 157
4.1 Best Practices in Naming Conventions 157
4.2 Best Practices in Database Design 158
5. Data Model Governance 158
5.1 Data Model and Design Quality Management 158
5.2 Data Modeling Metrics 160
6. Works Cited / Recommended 162

Chapter 6: Data Storage and Operations 165
1. Introduction 165
1.1 Business Drivers 167
1.2 Goals and Principles 167
1.3 Essential Concepts 168
2. Activities 188
2.1 Manage Database Technology 188
2.2 Manage Databases 190
3. Tools 202
3.1 Data Modeling Tools 202
3.2 Database Monitoring Tools 202
3.3 Database Management Tools 202
3.4 Developer Support Tools 203
4. Techniques 203
4.1 Test in Lower Environments 203
4.2 Physical Naming Standards 203
4.3 Script Usage for All Changes 203
5. Implementation Guidelines 204
5.1 Readiness Assessment / Risk Assessment 204
5.2 Organization and Cultural Change 204
6. Data Storage and Operations Governance 205
6.1 Metrics 205
6.2 Information Asset Tracking 206
6.3 Data Audits and Data Validation 206
7. Works Cited / Recommended 207

Chapter 7: Data Security 209
1. Introduction 209
1.1 Business Drivers 211
1.2 Goals and Principles 214
1.3 Essential Concepts 214

2. Activities 235
2.1 Identify Data Security Requirements 235
2.2 Define Data Security Policy 237
2.3 Define Data Security Standards 238
3. Tools 246
3.1 Anti-Virus Software / Security Software 246
3.2 HTTPS 246
3.3 Identity Management Technology 246
3.4 Intrusion Detection and Prevention Software 246
3.5 Firewalls (Prevention) 247
3.6 Metadata Tracking 247
3.7 Data Masking/Encryption 247
4. Techniques 247
4.1 CRUD Matrix Usage 247
4.2 Immediate Security Patch Deployment 247
4.3 Data Security Attributes in Metadata 248
4.4 Metrics 248
4.5 Security Needs in Project Requirements 250
4.6 Efficient Search of Encrypted Data 250
4.7 Document Sanitization 251
5. Implementation Guidelines 251
5.1 Readiness Assessment / Risk Assessment 251
5.2 Organization and Cultural Change 252
5.3 Visibility into User Data Entitlement 252
5.4 Data Security in an Outsourced World 252
5.5 Data Security in Cloud Environments 253
6. Data Security Governance 254
6.1 Data Security and Enterprise Architecture 254
7. Works Cited / Recommended 255

Chapter 8: Data Integration and Interoperability 257
1. Introduction 257
1.1 Business Drivers 258
1.2 Goals and Principles 260
1.3 Essential Concepts 260
2. Data Integration Activities 273
2.1 Plan and Analyze 273
2.2 Design Data Integration Solutions 276
2.3 Develop Data Integration Solutions 278
2.4 Implement and Monitor 280
3. Tools 280
3.1 Data Transformation Engine/ETL Tool 280
3.2 Data Virtualization Server 281
3.3 Enterprise Service Bus 281
3.4 Business Rules Engine 281
3.5 Data and Process Modeling Tools 281
3.6 Data Profiling Tool 282
3.7 Metadata Repository 282
4. Techniques 282

5. Implementation Guidelines 282
5.1 Readiness Assessment / Risk Assessment 282
5.2 Organization and Cultural Change 283
6. DII Governance 283
6.1 Data Sharing Agreements 284
6.2 DII and Data Lineage 284
6.3 Data Integration Metrics 285
7. Works Cited / Recommended 285

Chapter 9: Document and Content Management 287
1. Introduction 287
1.1 Business Drivers 289
1.2 Goals and Principles 289
1.3 Essential Concepts 290
2. Activities 305
2.1 Plan for Lifecycle Management 305
2.2 Manage the Lifecycle 308
2.3 Publish and Deliver Content 311
3. Tools 312
3.1 Enterprise Content Management Systems 312
3.2 Collaboration Tools 315
3.3 Controlled Vocabulary and Metadata Tools 315
3.4 Standard Markup and Exchange Formats 315
3.5 E-discovery Technology 317
4. Techniques 318
4.1 Litigation Response Playbook 318
4.2 Litigation Response Data Map 318
5. Implementation Guidelines 319
5.1 Readiness Assessment / Risk Assessment 319
5.2 Organization and Cultural Change 321
6. Documents and Content Governance 322
6.1 Information Governance Frameworks 322
6.2 Proliferation of Information 323
6.3 Govern for Quality Content 323
6.4 Metrics 324
7. Works Cited / Recommended 326

Chapter 10: Reference and Master Data 327
1. Introduction 327
1.1 Business Drivers 328
1.2 Goals and Principles 329
1.3 Essential Concepts 330
2. Activities 349
2.1 MDM Activities 349
2.2 Reference Data Activities 351
3. Tools and Techniques 353
4. Implementation Guidelines 353
4.1 Adhere to Master Data Architecture 354
4.2 Monitor Data Movement 354

4.3 Manage Reference Data Change 354
4.4 Data Sharing Agreements 355
5. Organization and Cultural Change 355
6. Reference and Master Data Governance 356
6.1 Metrics 356
7. Works Cited / Recommended 357

Chapter 11: Data Warehousing and Business Intelligence 359
1. Introduction 359
1.1 Business Drivers 361
1.2 Goals and Principles 361
1.3 Essential Concepts 362
2. Activities 372
2.1 Understand Requirements 372
2.2 Define and Maintain the DW/BI Architecture 372
2.3 Develop the Data Warehouse and Data Marts 373
2.4 Populate the Data Warehouse 374
2.5 Implement the Business Intelligence Portfolio 375
2.6 Maintain Data Products 376
3. Tools 378
3.1 Metadata Repository 379
3.2 Data Integration Tools 379
3.3 Business Intelligence Tools Types 380
4. Techniques 383
4.1 Prototypes to Drive Requirements 383
4.2 Self-Service BI 384
4.3 Audit Data that can be Queried 384
5. Implementation Guidelines 384
5.1 Readiness Assessment / Risk Assessment 385
5.2 Release Roadmap 385
5.3 Configuration Management 385
5.4 Organization and Cultural Change 386
6. DW/BI Governance 387
6.1 Enabling Business Acceptance 387
6.2 Customer / User Satisfaction 388
6.3 Service Level Agreements 388
6.4 Reporting Strategy 388
6.5 Metrics 389
7. Works Cited / Recommended 390

Chapter 12: Metadata Management 393
1. Introduction 393
1.1 Business Drivers 395
1.2 Goals and Principles 396
1.3 Essential Concepts 397
2. Activities 409
2.1 Define Metadata Strategy 409
2.2 Understand Metadata Requirements 410
2.3 Define Metadata Architecture 410

2.4 Create and Maintain Metadata 412
2.5 Query, Report, and Analyze Metadata 414
3. Tools 414
3.1 Metadata Repository Management Tools 414
4. Techniques 415
4.1 Data Lineage and Impact Analysis 415
4.2 Metadata for Big Data Ingest 417
5. Implementation Guidelines 417
5.1 Readiness Assessment / Risk Assessment 418
5.2 Organizational and Cultural Change 418
6. Metadata Governance 419
6.1 Process Controls 419
6.2 Documentation of Metadata Solutions 419
6.3 Metadata Standards and Guidelines 420
6.4 Metrics 420
7. Works Cited / Recommended 421

Chapter 13: Data Quality 423
1. Introduction 423
1.1 Business Drivers 426
1.2 Goals and Principles 426
1.3 Essential Concepts 427
2. Activities 446
2.1 Define High Quality Data 446
2.2 Define a Data Quality Strategy 447
2.3 Identify Critical Data and Business Rules 447
2.4 Perform an Initial Data Quality Assessment 448
2.5 Identify and Prioritize Potential Improvements 449
2.6 Define Goals for Data Quality Improvement 449
2.7 Develop and Deploy Data Quality Operations 450
3. Tools 457
3.1 Data Profiling Tools 457
3.2 Data Querying Tools 457
3.3 Modeling and ETL Tools 457
3.4 Data Quality Rule Templates 458
3.5 Metadata Repositories 458
4. Techniques 458
4.1 Preventive Actions 458
4.2 Corrective Actions 459
4.3 Quality Check and Audit Code Modules 460
4.4 Effective Data Quality Metrics 460
4.5 Statistical Process Control 461
4.6 Root Cause Analysis 462
5. Implementation Guidelines 462
5.1 Readiness Assessment / Risk Assessment 463
5.2 Organization and Cultural Change 464
6. Data Quality and Data Governance 465
6.1 Data Quality Policy 466
6.2 Metrics 466

7. Works Cited / Recommended 467

Chapter 14: Big Data and Data Science 469
1. Introduction 469
1.1 Business Drivers 470
1.2 Principles 471
1.3 Essential Concepts 472
2. Activities 482
2.1 Define Big Data Strategy and Business Needs 482
2.2 Choose Data Sources 483
2.3 Acquire and Ingest Data Sources 484
2.4 Develop Data Hypotheses and Methods 485
2.5 Integrate / Align Data for Analysis 485
2.6 Explore Data Using Models 485
2.7 Deploy and Monitor 487
3. Tools 488
3.1 MPP Shared-nothing Technologies and Architecture 489
3.2 Distributed File-based Databases 490
3.3 In-database Algorithms 491
3.4 Big Data Cloud Solutions 491
3.5 Statistical Computing and Graphical Languages 491
3.6 Data Visualization Tools 491
4. Techniques 492
4.1 Analytic Modeling 492
4.2 Big Data Modeling 493
5. Implementation Guidelines 493
5.1 Strategy Alignment 494
5.2 Readiness Assessment / Risk Assessment 494
5.3 Organization and Cultural Change 495
6. Big Data and Data Science Governance 495
6.1 Visualization Channels Management 496
6.2 Data Science and Visualization Standards 496
6.3 Data Security 496
6.4 Metadata 497
6.5 Data Quality 497
6.6 Metrics 498
7. Works Cited / Recommended 499

Chapter 15: Data Management Maturity Assessment 501
1. Introduction 501
1.1 Business Drivers 503
1.2 Goals and Principles 503
1.3 Essential Concepts 504
2. Activities 508
2.1 Plan Assessment Activities 509
2.2 Perform Maturity Assessment 511
2.3 Interpret Results 512
2.4 Create a Targeted Program for Improvements 513
2.5 Re-assess Maturity 513

3. Tools 514
4. Techniques 514
4.1 Selecting a DMM Framework 514
4.2 DAMA-DMBOK Framework Use 515
5. Guidelines for a DMMA 515
5.1 Readiness Assessment / Risk Assessment 515
5.2 Organizational and Cultural Change 516
6. Maturity Management Governance 516
6.1 DMMA Process Oversight 517
6.2 Metrics 517
7. Works Cited / Recommended 518

Chapter 16: Data Management Organization and Role Expectations 519
1. Introduction 519
2. Understand Existing Organization and Cultural Norms 519
3. Data Management Organizational Constructs 521
3.1 Decentralized Operating Model 521
3.2 Network Operating Model 522
3.3 Centralized Operating Model 522
3.4 Hybrid Operating Model 523
3.5 Federated Operating Model 524
3.6 Identifying the Best Model for an Organization 525
3.7 DMO Alternatives and Design Considerations 525
4. Critical Success Factors 526
4.1 Executive Sponsorship 526
4.2 Clear Vision 527
4.3 Proactive Change Management 527
4.4 Leadership Alignment 527
4.5 Communication 527
4.6 Stakeholder Engagement 528
4.7 Orientation and Training 528
4.8 Adoption Measurement 528
4.9 Adherence to Guiding Principles 529
4.10 Evolution Not Revolution 529
5. Build the Data Management Organization 529
5.1 Identify Current Data Management Participants 529
5.2 Identify Committee Participants 530
5.3 Identify and Analyze Stakeholders 530
5.4 Involve the Stakeholders 531
6. Interactions Between the DMO and Other Data-oriented Bodies 531
6.1 The Chief Data Officer 531
6.2 Data Governance 532
6.3 Data Quality 533
6.4 Enterprise Architecture 533
6.5 Managing a Global Organization 534
7. Data Management Roles 534
7.1 Organizational Roles 535
7.2 Individual Roles 535
8. Works Cited / Recommended 537

Chapter 17: Data Management and Organizational Change Management 539
1. Introduction 539
2. Laws of Change 540
3. Not Managing a Change: Managing a Transition 540
4. Kotter's Eight Errors of Change Management 543
4.1 Error #1: Allowing Too Much Complacency 543
4.2 Error #2: Failing to Create a Sufficiently Powerful Guiding Coalition 544
4.3 Error #3: Underestimating the Power of Vision 544
4.4 Error #4: Under Communicating the Vision by a Factor of 10, 100, or 1000 545
4.5 Error #5: Permitting Obstacles to Block the Vision 545
4.6 Error #6: Failing to Create Short-Term Wins 545
4.7 Error #7: Declaring Victory Too Soon 546
4.8 Error # 8: Neglecting to Anchor Changes Firmly in the Corporate Culture 547
5. Kotter's Eight Stage Process for Major Change 547
5.1 Establishing a Sense of Urgency 548
5.2 The Guiding Coalition 551
5.3 Developing a Vision and Strategy 555
5.4 Communicating the Change Vision 558
6. The Formula for Change 562
7. Diffusion of Innovations and Sustaining Change 563
7.1 The Challenges to be Overcome as Innovations Spread 565
7.2 Key Elements in the Diffusion of Innovation 565
7.3 The Five Stages of Adoption 565
7.4 Factors Affecting Acceptance or Rejection of an Innovation or Change 566
8. Sustaining Change 567
8.1 Sense of Urgency / Dissatisfaction 568
8.2 Framing the Vision 568
8.3 The Guiding Coalition 568
8.4 Relative Advantage and Observability 568
9. Communicating Data Management Value 569
9.1 Communications Principles 569
9.2 Audience Evaluation and Preparation 570
9.3 The Human Element 571
9.4 Communication Plan 571
9.5 Keep Communicating 572
10. Works Cited / Recommended 573

Acknowledgements 575

Index 579