Use Best Practice Patterns to Understand and Architect Manageable, Efficient Information Supply Chains That Help You Leverage All Your Data and Knowledge

In the era of “Big Data,” information pervades every aspect of the organization. Therefore, architecting and managing it is a multi-disciplinary task. Now, two pioneering IBM® architects present proven architecture patterns that fully reflect this reality. Using their pattern language, you can accurately characterize the information issues associated with your own systems, and design solutions that succeed over both the short- and long-term.

Building on the analogy of a supply chain, Mandy Chessell and Harald C. Smith explain how information can be transformed, enriched, reconciled, redistributed, and utilized in even the most complex environments. Through a realistic, end-to-end case study, they help you blend overlapping information management, SOA, and BPM technologies that are often viewed as competitive.

Related Books of Interest

Data Integration Blueprint and Modeling
Techniques for a Scalable and Sustainable Architecture
By Anthony David Giordano

Making Data Integration Work: How to Systematically Reduce Cost, Improve Quality, and Enhance Effectiveness
As businesses increasingly rely on analytics, the need for a blueprint for data integration is increasing now more than ever. This book presents the solution: a clear, consistent approach to defining, designing, and building data integration components to reduce cost, simplify management, enhance quality, and improve effectiveness. Leading IBM data management expert Tony Giordano brings together best practices for architecture, design, and methodology, and shows how to do the disciplined work of getting data integration right.

Patterns of Information Management
By Mandy Chessell, Harald Smith

Use Best Practice Patterns to Understand and Architect Manageable, Efficient Information Supply Chains That Help You Leverage All Your Data and Knowledge
In the era of “Big Data,” information pervades every aspect of the organization. Therefore, architecting and managing it is a multi-disciplinary task. Now, two pioneering IBM® architects present proven architecture patterns that fully reflect this reality. Using their pattern language, you can accurately characterize the information issues associated with your own systems, and design solutions that succeed over both the short- and long-term.

Building on the analogy of a supply chain, Mandy Chessell and Harald C. Smith explain how information can be transformed, enriched, reconciled, redistributed, and utilized in even the most complex environments. Through a realistic, end-to-end case study, they help you blend overlapping information management, SOA, and BPM technologies that are often viewed as competitive.

Sign up for the monthly IBM Press newsletter at ibmpressbooks.com/newsletters
Enterprise Master Data Management
An SOA Approach to Managing Core Information
By Allen Dreibelbis, Eberhard Hechler, Ivan Milman, Martin Oberhofer, Paul Van Run, Dan Wolfson
The Only Complete Technical Primer for MDM Planners, Architects, and Implementers

Enterprise Master Data Management provides an authoritative, vendor-independent MDM technical reference for practitioners: architects, technical analysts, consultants, solution designers, and senior IT decision makers. Written by the IBM data management innovators who are pioneering MDM, this book systematically introduces MDM’s key concepts and technical themes, explains its business case, and illuminates how it interrelates with and enables SOA.

Drawing on their experience with cutting-edge projects, the authors introduce MDM patterns, blueprints, solutions, and best practices published nowhere else—everything you need to establish a consistent, manageable set of master data, and use it for competitive advantage.

Analytics Across the Enterprise
How IBM Realizes Business Value from Big Data and Analytics
By Brenda L. Dietrich, Emily C. Plachy, Maureen F. Norton
How to Transform Your Organization with Analytics: Insider Lessons from IBM’s Pioneering Experience

Analytics Across the Enterprise demystifies your analytics journey by showing you how IBM has successfully leveraged analytics across the enterprise, worldwide. Three of IBM’s pioneering analytics practitioners share invaluable real-world perspectives on what does and doesn’t work and how you can start or accelerate your own transformation. This book provides an essential framework for becoming a smarter enterprise and shows through 31 case studies how IBM has derived value from analytics throughout its business.

Whatever your industry or role, whether a current or future leader, analytics can make you smarter and more competitive. Analytics Across the Enterprise shows how IBM did it—and how you can, too.
Related Books of Interest

### DB2 Essentials
**Understanding DB2 in a Big Data World, 3rd Edition**
By Raul F. Chong, Clara Liul

*The Easy, Visual Introduction to IBM DB2 Version 10.5 for Linux, UNIX, and Windows*

*DB2 Essentials* covers everything you need to get productive with the latest version of IBM DB2 and apply it to today’s business challenges. It discusses key features introduced in DB2 Versions 10.5, 10.1, and 9.7, including improvements in manageability, integration, security, Big Data support, BLU Acceleration, and cloud computing.

Raul F. Chong and Clara Liu illuminate key concepts with examples drawn from their extensive experience with DB2 in enterprise environments. They explain how DB2 has evolved, what’s new, and how to choose the right products, editions, and tools and then walk you through installation, configuration, security, data access, remote connectivity, and day-to-day administration.

---

### IBM Cognos Business Intelligence v10
*The Complete Guide*
By Sangeeta Gautam

### Decision Management Systems
*A Practical Guide to Using Business Rules and Predictive Analytics*
By James Taylor

### The New Era of Enterprise Business Intelligence
*Using Analytics to Achieve a Global Competitive Advantage*
By Mike Biere

### An Introduction to IMS
*Your Complete Guide to IBM Information Management System*
By Barbara Klein, Richard Alan Long, Kenneth Ray Blackman, Diane Lynne Goff, Stephen Paul Nathan, Moira McFadden Lanyi, Margaret M. Wilson, John Butterweck, Sandra L. Sherrill

### IBM Cognos 10 Report Studio
*Practical Examples*
By Filip Draskovic, Roger Johnson

---

Sign up for the monthly IBM Press newsletter at ibmpressbooks.com/newsletters
This page intentionally left blank
Performing Information Governance
This page intentionally left blank
The following terms are trademarks of International Business Machines Corporation in many jurisdictions worldwide: IBM, IBM Press, Cognos, MQSeries, InfoSphere, DataStage, and TM1. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at “Copyright and trademark information” at www.ibm.com/legal/copytrade.shtml.

Microsoft, PowerPoint, and Excel are trademarks of Microsoft Corporation in the United States, other countries, or both. Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates. Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both. UNIX is a registered trademark of The Open Group in the United States and other countries. Other company, product, or service names may be trademarks or service marks of others.

Library of Congress Control Number: 2014943598

All rights reserved. This publication is protected by copyright, and permission must be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or likewise. To obtain permission to use material from this work, please submit a written request to Pearson Education, Inc., Permissions Department, One Lake Street, Upper Saddle River, New Jersey 07458, or you may fax your request to (201) 236-3290.


Text printed in the United States on recycled paper at R.R. Donnelley in Crawfordsville, Indiana.

First printing: September 2014
Contents

Preface ......................................................... xix
Acknowledgments ................................. xxii
About the Author ................................. xxiv

Introduction: Why Another Book on Information Governance? . . . .1
What Is Not Working in Information Governance .............................................. 2
Providing the “How” of Information Governance ................................................. 3

Part I: Overview of Information Governance & Enterprise Information Management

Chapter 1 Information Governance & EIM Overview Governance ................................. 7
What Is Enterprise Information Management? ...................................................... 7
The Functions of EIM .................................................. 8
   Data Management: EIM’s Technical Development and Management Discipline ........ 9
What Is Information Governance? ................................................................. 10
The Information Governance Component Framework ...................................... 10
   Information Governance Organization Component ...................................... 11
   Data Stewardship Component ................................................................. 14
   Data Quality Management Component .................................................. 17
   Metadata Management Component ........................................................ 23
   Structural Metadata .............................................................................. 26
   Navigational Metadata ........................................................................ 27
   Privacy and Security Component ............................................................ 32
   Information Life Cycle Management Component ..................................... 34
Information Governance in the Context of EIM ............................................... 35
Summary ............................................................. 36
## Chapter 2  Other Core EIM Functions

End-of-Chapter Review Questions .......................................................... 37
Essay Question .................................................................................. 37
For Further Reading ....................................................................... 37

### Transaction Processing Function
Business 101: Reviewing the Structure and Automation of the Business Model 40
Organizational Structure ................................................................. 40

### MDM Function
What Is Master Data? ....................................................................... 47
What Is Master Data Management? ................................................... 48
MDM Architectural Patterns ............................................................ 49
Operational and Analytic Master Data Management Implementations 50

### BI Function
Analytic Architectural Pattern ......................................................... 51
Data Warehouse Architectural Pattern ............................................. 53
Information Governance Considerations ......................................... 55
Data Integration Architectural Pattern ............................................. 57

### Big Data
An EIM View .................................................................................. 60

### Summary
End-of-Chapter Review Questions .................................................... 64
Essay Question ................................................................................ 65

## Part II  Performing Information Governance Activities in EIM Projects

### Chapter 3  Preparing the Information Governance Organization

End-of-Chapter Review Questions ..................................................... 69
Essay Question ................................................................................ 69
For Further Reading ....................................................................... 69
Chapter 4 Performing Information Governance Tasks in Transactional Projects

Performing IG on Custom Transaction Processing Projects
What Is the Agile Method?
Agile Custom Transaction Processing SDLC with Information Governance Activities
1. Iteration 0: Initiate Phase
2. Construction Iteration
3. Transition Iteration
Agile Custom Transaction Processing SDLC Method Review
Case Study: Applying the Agile SDLC with Information Governance Activities for a Claims Reengineering Project
Initiate Project Planning Considerations
Initiate Phase
Performing Information Governance on COTS Transaction Processing Projects
Information Governance Implications with COTS Packages
COTS Transaction Processing SDLC with Information Governance Activities
1. Business Blueprint Phase
2. Mapping and Construction Phase
COTS Transaction Processing SDLC Method Review
Case Study: Applying the COTS SDLC with IG Activities for a Finance Consolidation Project
Pharma Co’s Finance Business Blueprint Phase Project Planning Considerations
Sample Pharmaceutical Finance Business Blueprint Phase
Summary
End-of-Chapter Review Questions
Essay Question

Chapter 5 Performing Information Governance Tasks in Master Data Management Projects

The Interrelationship of Master Data Management and Information Governance
Performing Information Governance on MDM Projects
Master Data Management Architectural Components
Prioritizing Master Data Management Project Iterations
Master Data Management Systems Development Life Cycle
MDM SDLC with Information Governance Activities
1. Master Data Management Strategy Phase
2. Master Data Management Architectural Blueprint Phase
Chapter 8  Performing Information Governance Tasks in “Big Data” Projects .............................. 329

A Review of Big Data ........................................................................................................ 329

Big Data Discovery ........................................................................................................... 330

The Role of the Data Steward in Big Data Projects ......................................................... 331

The Big Data Paradox of Data Ownership ...................................................................... 333

Big Data Control ............................................................................................................... 333

Controlling Big Data Through the Intake Processes ....................................................... 334

Big Data Usage ................................................................................................................ 339

Possible Big Data Usage Opportunities ........................................................................ 339

New Patterns of Big Data Usage .................................................................................. 340

A Life Cycle for Big Data ............................................................................................... 343

The Big Data SLC ............................................................................................................ 344

The Big Data Project Data Steward .............................................................................. 345

1.  Big Data Discovery Phase ......................................................................................... 345

2.  Big Data Control Phase ......................................................................................... 350

3.  Big Data Usage Phase ........................................................................................... 355

Big Data Systems Life Cycle Method Review ................................................................. 362

Case Study: Leveraging Big Data for a Pharmaceutical Clinical Trial SLC .................. 362

2.  Clinical Research Big Data Control Phase ............................................................... 363

3.  Clinical Research Big Data Usage Phase ................................................................. 366

Summary ......................................................................................................................... 366

Part II Review: Performing Information Governance Activities in Enterprise Information
Management Projects ........................................................................................................ 367

End-of-Chapter Review Questions .................................................................................. 367

Essay Question ............................................................................................................... 367
Part III: Performing Ongoing Information Governance Processes

Chapter 9  Information Governance Organization Operations . .371

- Internal Information Governance Council Operations . .372
- Information Governance Leaders . .373
- Information Governance Organization Stakeholders . .374

Information Governance Organization Operations . .376

- Data Stewardship Community Operations . .376
- The Data Quality Organization Operations . .377
- Managing the Metadata Function . .379
- Managing the Information Security Function . .381

The Expanding Role of Information Governance . .382

Managing Enterprise Data Management . .383

- Integrating Master Data Management into the Information Governance Organization . .386
- Integrating Business Intelligence into the Information Governance Organization . .387

- Integrating Application Development into the Information Governance Organization . .389

The Operational Aspect of Information Governance Policies . .390

- Creating Information Governance Policies . .391
- Creating or Enhancing Information Governance Policies . .392

Information Governance Policy Enforcement Models . .396

Case Study: Modifying the Healthcare Company’s Information Governance Organization for HIPAA . .397

- Detailed Breakdown of the Information Governance Policy Change Activities . .398

Summary . .399

End-of-Chapter Review Questions . .399

Essay Question . .399

Chapter 10  Ongoing Data Stewardship Processes . .401

Overview of Data Stewardship . .401

- The Mission of a Data Steward . .402
- The Structure of Data Stewardship Community . .403

The Expanding Role of Data Stewardship in Project Work . .405

- Data Stewardship in General Project Work . .405
- Data Stewardship in Custom Development Transactional Project Work . .407
- Data Stewardship in Master Data Management Project Work . .409
- Data Stewardship in Business Intelligence Project Work . .409

Ongoing Data Stewardship Operations . .411

- Data Ownership Administration . .411
- Data Stewardship Issue-Resolution Activities . .413
## Data Stewardship Data Quality Management Activities

- Data Stewardship Data Quality Management Activities: 415
- Information Life Cycle Management Audits: 420
- Information Security Audits: 421
- Operational Data Stewardship Within the Business Community: 423

## Chapter 11 Ongoing Data Quality Management Processes

**Review of the Data Quality Management Processes:**
- What Is Data Quality?: 430
- Review of the Data Quality Framework: 430

**Data Quality Life Cycle:**
- The Define Phase: 436
- The Audit Phase: 445
- The Improve Phase: 452

**Data Quality Remediation Life Cycle:**
1. Data Quality Remediation Solution Outline Activity: 454
2. Data Quality Remediation Design Activity: 457
3. Data Quality Remediation Build Cycle Activity: 459
4. Data Quality Remediation Deployment Activity: 460

Case Study: The Healthcare Company’s Data Quality ICD-10 Compliance
1. ICD-10 Data Quality Remediation Solution Outline Activity: 462
2. ICD-10 Data Quality Remediation Design Activity: 464
3. ICD-10 Data Quality Remediation Build Cycle Activity: 465
4. ICD-10 Data Quality Remediation Deployment Activity: 466

**Summary:** 466

**End-of-Chapter Review Questions:** 466

**Essay Question:** 467

## Chapter 12 Ongoing Metadata Management Processes

**Overview of Metadata:** 469

**Metadata Management:** 470

**Categories of Metadata:** 472

**Users of Metadata:** 478

**The Evolution of Metadata:** 478

**The Growing Importance of Metadata:** 479

**Metadata Management Administration in Information Governance:** 480

**Metadata Capture Phase:** 481

**Metadata Usage Phase:** 485

**Metadata Maintenance Phase:** 487
2. Master Data Management SDLC with Information Governance Activities .................. 579  
   2.1 Case Study 3: Pharmaceutical Customer MDM Project ......................... 586  
3. Business Intelligence SDLC with Information Governance Activities ................. 590  
   3.1 Case Study 4: A BI Environment for the Healthcare Company ....................... 598  
4. Information Lifecycle Management (ILM) SDLCs ........................................ 610  
   4.1 Proactive ILM SDLC ................................................................. 611  
   4.2 ILM Renovation Project SDLC ...................................................... 613  
   4.3 Case Study 5: The Healthcare Company Legacy Claims Systems Renovation. .................................................. 614  
5. Big Data SDLC ..................................................................................... 616  
   5.1 Big Data SLC ................................................................................. 616  
   5.2 Case Study 6L The Pharmaceutical Company Clinical Research Big Data Project .... 617  

Appendix D  Glossary of Information Governance Terms ....... 619  

Appendix E  Case Study Models ....................................................... 623  

Index ................................................................. 627
This text provides a practitioner’s guide on how to apply information governance principles and processes in both projects and ongoing enterprise information management (EIM) operations. Information governance has been recognized as a critical organizational process for both information management best practices and regulatory compliance. However, of all the EIM functions, information governance cannot be performed successfully as a standalone discipline. Information governance when practiced in a vacuum has not brought the value expected and intended. Although there are many sources for why information governance is important, few go into detail as to how to, step by step, execute information governance processes in both development projects and in ongoing organization processes.

This book introduces you to the core components of information governance and how they “thread” into the various functions of EIM. It also covers in detail how to pragmatically and practically execute information governance functions on development projects and in on-going organizational processes.

**Intended Audience**

This text serves many different audiences. Experienced data management professionals can use it to confirm the activities, tasks, and best practices for performing information governance. College students can use this book as a textbook in an upper-level information management college curriculum. The intended audience includes the following:

- Chief information officers
- Chief data officers
- Business and technical data stewards
- Data quality analysts and auditors
- Metadata management professionals
- Master data management professionals
• Program/project managers
• Data modeling and database practitioners
• Other information management practitioners
• Information management-focused college students

Scope of the Text
This book stresses the implementation and operational aspects of implementing information governance using the six core components of information governance and how they thread into other EIM functions such as transactional processing, business intelligence (BI), and master data management (MDM).
With that goal in mind, this book:
• Reviews the functions of EIM and the components of information governance
• Provides a step-by-step approach to performing project-level information governance activities within each of the EIM functions
• Provides a step-by-step approach to ongoing information governance activities within the organization
• Reviews case studies for each of the project-level and operational information governance activities

Organization of the Text
The text is organized into three parts:
• Part I: Overview of Information Governance & Enterprise Information Management
   The first part of this text provides an overview of EIM. Because of the broad nature of information governance and its potential implications in all facets of information management, it is important to understand the core functions within the scope of EIM, such as transactional processing, BI, and MDM, as well as the emergence of “big data.” Then to clarify how information governance is a “horizontal thread” to the other EIM functions, the core components of information governance are defined.
Part II: Performing Information Governance Activities in Enterprise Information Management Projects

The second part of the text covers the systems development life cycle (SDLC) activities of an information governance project in terms of the activities, tasks, and deliverables with each of the three EIM functions. For example, it covers the data stewardship, metadata management, and data quality tasks in the definition, design, and deployment of a data warehouse within a BI environment. It also covers transactional processing and MDM. It concludes with thoughts on big data and the information life cycle management (ILM).

Part III: Performing Ongoing Information Governance Processes

The third part of this text starts with a discussion on the evolution of information governance organization models, optimal models for how to align information governance organizations within the broader information, and the typical roles in an information governance organization. It then covers the roles of data stewardship with business and technical organizations. Next, it reviews ongoing data quality processes in terms of the types of ongoing data quality projects. It concludes by reviewing ongoing metadata management functions.

Also in this text, each chapter provides a set of questions on the core concepts in the book to test your comprehension of the material. You can find answers to the questions for each chapter in Appendix B, “Chapter Exercise Answers.”

Much of the supporting materials to the text can be found in the appendices, which include the following:

- Appendix A, “Organizational Change Management in Information Governance”—This appendix discusses techniques on how to overcome the organizational challenges of implementing an information governance process and organization.

- Appendix B, “Chapter Exercise Answers”—This appendix contains answers to the questions found at the end of each chapter.

- Appendix C, “Enterprise Information Management Systems Development Life Cycles”—This contains the SDLCs and the example of how they have been applied from the chapters.

- Appendix D, “Glossary of Information Governance Terms”—This appendix contains the glossary of terms used in the book.

- Appendix E, “Case Study Models”—This appendix reviews the information governance case studies found throughout the text.
Acknowledgments

Information governance is a broad and pervasive information management function with many perspectives and points of view. The processes defined in this book are a result of actual project work that is a product of countless hours defining processes, testing processes, and using these processes in many organizations. These efforts can only be performed in collaboration with knowledgeable, dedicated, and experience practitioners. In particular, I want to acknowledge Glenn Finch, Sharon Hodgson, Ron Shelby, Linc Markham, Brian Scheld, Dan Sutherland, and Brett Gow, all of whom played an integral part in the development of this book.
Dedication

I want to dedicate this book to my daughters, Katie and Kelsie; they are my greatest joy and blessing.

Also, I have to thank my two Irish wolfhounds, Rupert and Switters, who kept me company during the long hours writing this book. They are truly man’s best friends.
About the Author

Anthony Giordano is an experienced executive with more than 25 years of global professional services experience in the information management field in the areas of business intelligence, data warehousing, customer relationship management, and program management. He has significant work experience in the financial services, life sciences, and automotive sectors. Mr. Giordano has functioned in various roles, including positions such as IBM Japan BAO Service Line Leader, Financial Services Practice Partner, and other consulting roles. He has extensive international experience in Australia, Thailand, Turkey, England, and living in Japan. Tony has been dedicated to implementing customer-focused business analytics and information management platform that deliver business outcomes.

In his role at Merkle, Tony is responsible for the Technology Solutions Group, which defines designs, develops, and deploys the Foundation Marketing Platform’s that Merkle’s customer run their multi-channel campaigns and marketing analytics.

Tony joins Merkle from IBM’s Business Analytics and Optimization Practice, where he spent 18 years in a variety of senior-level positions. Most recently, Tony led IBM’s BAO Japan Service Line. He also held IBM’s BAO Global Leader for Banking and Financial Markets, Enterprise Information Management Service Line for North America, and Financial Services BAO Partner roles. Prior to IBM and PricewaterhouseCooper’s, Tony held consulting roles in Oracle.

Tony holds a Master of Business Administration degree from Wayne State University in Detroit, majoring in accounting, and a Bachelor of Business from Walsh College, majoring in Computer Information Systems.

Tony is the author of Data Integration Blueprint and Modeling (IBM Press).
Introduction: Why Another Book on Information Governance?

Information governance is not a new discipline; in fact, it has been an acknowledged practice going back to the 1980s. Every 2 to 3 years, there is a renewed interest in information governance due to a new technology or set of government regulations. It has evolved in the past 20 years from multiple areas of the business enterprise. For example, data stewardship and data quality evolved from data management in areas such as data modeling, where the need to define the business and technical metadata is required to understand the optimal design blueprint for the target data model.

The need to redefine how the enterprise’s data is protected is a result of the ever-increasing blur between the enterprise and the external world. This has developed many of the security and privacy aspects of information governance—resulting in a broad discipline with many interfaces, control points, and processes. Figure I.1 shows how this involves both business and technology.
What Is Not Working in Information Governance

Many enterprises have started programs or projects covering all or some aspects of information governance, but many others have also abandoned their efforts, frustrated with their inability to sustain the implementation and achieve results. Enterprises fail in their information governance efforts for many reasons, including a lack of

- Clarity of purpose—Many organizations have bought into the “what” of information governance but have struggled to understand the “why.” Success, in some companies, may be defined as implementing information governance processes. Here, companies may struggle to rigorously implement every component of a process without recognizing how information governance processes help the organization achieve operational and performance benefits. In these cases, implementing a process is less the measure of success than is the achievement of well-defined and well-understood performance benefits of that process improvement.
• **Organizational commitment**—Some organizations believe that they need to have an information governance process or organization to meet regulatory requirements or government mandates. After fulfilling that obligation, they perceive that the need for continuing information governance no longer exists and so they discontinue/disband the process/organization. Other organizations face a change in leadership, where the new leadership has not bought into the value of having an information governance process. Or in other words, these organizations never bought into the sustained business value of information governance (the “why” of information governance).

• **Political tradeoffs and organizational change management**—Developing an information capability and competency requires the balancing of current practices against practices optimized for enterprise performance. This often requires consideration of mid- to longer-term adjustments that, although less efficient in the near term, produce more consistent and predictable results. Some of these adjustments include adjusting to new roles, learning new processes, and adapting to broader and more integrated definitions of performance. New decision-making processes are some of the more challenging adjustments in that decisions benefiting individuals and specific lines of business or functional areas are balanced against broader enterprise performance requirements.

• **Understanding how to design and coordinate enterprise information governance**—These enterprises have bought into the “why” and “how” of information governance. They have bought the books, went to the seminars, brought in the “experts” and have sold information governance to their management, staff, and stakeholders, and may have even instantiated an organization. However, they fail due to a lack of actual performance. They simply do not integrate their information governance activities within development projects, and ongoing activities create additional project and process overhead. They are invariably shut down from a cost/benefit perspective.

**Providing the “How” of Information Governance**

The scope of this text is not on the “why” and “what” of information governance. This book provides prescriptive instructions, a step-by-step “how” to perform information governance. It discusses how to build activities on ongoing enterprise information management projects and ongoing processes, as portrayed in Figure I.2.
I. “Why” Information Governance?

- Lost productivity due to reconciling, correcting, rationalizing, combining data
- Financial restatements due to incorrect definitions of aggregated data
- Poor data quality costs
- Fraud due to inadequate privacy and security controls

II. “What” Is Information Governance?

- Information governance organization component
- Data stewardship component
- Data quality management component
- Metadata management component
- Information governance change management component
- Privacy and security component
- Information life cycle management component

III. “How” to Implement & Perform Information Governance

- Performing information governance tasks in business intelligence
- Performing information governance tasks in master data management
- Performing information governance tasks in transactional projects
- Performing information governance tasks in information life cycle management projects
- Performing information governance tasks in “big data” projects
- Ongoing data stewardship processes
- Ongoing data quality processes
- Ongoing metadata management processes

Figure I.2  The focus on performing information governance

It covers how to integrate information governance within existing efforts rather than create duplicative activities and tasks. Duplication unnecessarily burdens projects and creates a negative environment for information governance. Instead, this book places key information governance tasks and roles with the systems development life cycle of the various enterprise information management disciplines and facilitates and hardens the development of the various information management components.

After an information governance organization has been created, the text covers how to practically sustain and audit the information environment in such a way to ensure the highest ongoing data quality, data security, and data understanding.
This page intentionally left blank
Part I of this text provides an overview of enterprise information management (EIM) and information governance. It reviews and details the various functions in EIM, such as business intelligence, master data management, and transactional data management, with a discussion on the emerging function of “big data.” It next introduces and details the seven components of information governance. Although this is not a “what” is information governance book, it is necessary to establish a foundation for what each component is and how to deploy them in both development and ongoing efforts.

Part I sets the stage for the major focus of the book: how information governance components thread into EIM projects. For example, what data stewardship tasks are necessary when building a data mart data model? What types of metadata are captured and managed in a master data management project? What are the ongoing privacy and security tasks needed for connecting your intranet to a social network? These are presented at a high level and then defined and explained at a detailed level in subsequent sections and chapters of the book.

This chapter focuses on each of the functions within EIM in terms of their purpose, unique artifacts that are used to design and maintain the applications, and how they are built. It explains how these different functions interconnect to provide an entire information environment for an organization.

What Is Enterprise Information Management?

The major focus of this book is performing information governance within the context of enterprise information management (EIM.) To understand information governance, it is important to start with an understanding of EIM area within an organization. EIM is the information management foundation for both transaction and analytic processing. It is a series of functions that are segmented based on the type of transactional, operational, and analytic processes they support (see Figure 1.1).
The Functions of EIM

These EIM functions have similar design patterns around data and provide the context for process areas such as sales and marketing, finance, and production. For the purpose of this text, each is defined as follows:

- **The transaction processing function**—Centers on the creation and maintenance of the core business transactions in the business. This function is still 60% to 70% of all information technology (IT) budgets and defines what the transactions mean and how it is used within the organization.

- **The master data management (MDM) function**—Concentrates on the creation and maintenance of the core domain definitional information of an organization. It provides the context for our transactional and analytic data. For example, it provides the definition of what a customer is and what a product is. These definition and instantiated data elements are used in creating transactions and determining the measures needed to analyze what is a customer or how much of a product is used.

- **The business intelligence (BI) function**—Focuses use of data for different types of information analysis. A BI environment is the most data-centric of all EIM functions. It captures, collates, and conforms data from many disparate sources into a set of repositories in various structures for the many different types of reporting, descriptive, and predictive analytics used by disparate end users. A BI environment now offers their organizations a centralized environment to provide financial and marketing reporting and analytics.

Other authors and organizations may have different perspectives of what EIM consists of and the functional processes that it covers. This book, though, focuses on how to perform
information governance activities and tasks within the development and ongoing operations in these three EIM functions.

Data Management: EIM’s Technical Development and Management Discipline

To understand how information governance interacts in EIM functions, it is important to understand how EIM functions are developed and maintained. This section discusses the technical discipline of data management. Common patterns exist in the data-driven aspects of the three EIM functions. They have similar requirements and patterns in the blueprints, development life cycles, and maintenance of the applications. Over the past 30 years, IT has evolved a technical discipline known as data management.

Data management is the development and maintenance of architectures, best practices, and procedures that manage the full data life cycle of an organization. It is within data management that data architecture artifacts such as data models, data integration models, and information access patterns are developed and maintained.

The best example of a well-known data management process is data modeling. The systems development life cycle (SDLC) details how data models capture business requirements of an organization. It determines how to best structure those requirements into the different types of technical structures that are available: transactional, operational, and analytic (data warehouse, dimensional), as shown in Figure 1.2.

![Data Model SDLC Tasks](image)

**Figure 1.2** Data modeling in SDLC and maintenance tasks
The Relationship Between Data Management and Information Governance

A very tight relationship exists between data management and information governance. Often, the two are confused as the same discipline or overlap in areas such as metadata and data quality management. Within the development of the data management artifacts such as data models, there are information governance tasks such as business definitions of the entities, attributes, and relationships. Chapter 4, “Performing Information Governance Tasks in Transactional Projects,” explores these relationships in much greater detail. For this section, you just need to understand that a data management artifact is the blueprint for a database or data integration process and that the information governance aspects give it business context.

What Is Information Governance?

There are many definitions and points of view on what information governance is and what it is not. For this book, the formal definition of information governance is as follows:

> Information governance is the orchestration of people, process, and technology to enable an organization to leverage data as an enterprise asset.

Although information governance spans both business and technology (as shown in Figure I.1 in the Introduction), it is truly a business function with its primary directive to establish the policies for the creation and usage of data with an organization. It is an integral aspect of the understanding of an organization, which leads to the position that information governance should be considered an ongoing organizational function on par with accounting or marketing.

The Information Governance Component Framework

Information governance is a broad discipline that encompasses the definition, creation, use, security, ownership, and deletion of all organizational data. The information governance component framework covers the organizational models and roles needed to define and manage the policies and processes that affect the business and technical creation, maintenance, and usage of data within the enterprise. These components of information governance include the following:

- Information governance organization component
- Data stewardship component
- Data quality management component
- Metadata management component
- Privacy and security component
- Information life cycle management component
Information Governance Organization Component

The information governance organizational component is the “people” aspect of the discipline. It sets the policies for information governance and maintains a staff to support those policies in managing the development and ongoing usage of corporate information. Because information governance is an organizational process similar to accounting or marketing, as mentioned earlier, it requires a staffing structure capable of performing both project and ongoing activities and tasks and fitting into the broader organization. Many information governance programs and organizations fail because of their inability to work within the corporate culture and, when necessary, modify that corporate culture. It also depends on how the organization is organized and how successful its reporting chain is.

The Information Governance Organizational Model

There are several information governance organizational models, all of which revolve around certain key functions, including the information governance council and the data stewardship community function.

The information governance council (IGC) function focuses on setting the vision and goals, providing alignment within the broader organization, and setting the direction for the information governance process. The IGC function includes establishing the policies and procedures for information governance, such as the following:

- Data as a corporate asset policy
- Data creation and usage policy
- Data security requirements policies
- Data regulatory requirements (e.g., Sarbanes-Oxley) policies
- Data quality audit policies

The IGC is traditionally organized with key stakeholders from core functional areas such as accounting, marketing, research and development, sales, and production. Often, the department leaders will chair the IGC or provide an advocate. The IGC is traditionally led by a chief data officer (CDO).

The CDO has been an evolving role over the past 15 years and originally was only responsible for the information governance organization in terms of overall direction and day-to-day oversight. In recent years, many organizations have been expanding the responsibilities of the CDO to include oversight and day-to-day management of data development and maintenance functions. The CDO role and responsibility often includes the following:

- Owning and driving the organization’s data strategy and enterprise-level data vision
- “Selling” information governance, by driving data ownership and accountability in the business
- Directing data quality practices across the organization
• Aligning business and IT to support data quality through consistent business definitions and well-defined data quality ranges in transactional and analytic applications
• Providing leadership by sitting on executive committees where data programs and projects are approved and sponsored to ensure information governance processes are embedded into those programs
• Working with other business executives to understand their data quality requirements, objectives, and issues
• Providing leadership and support to members of the data stewardship community as they define data and metadata
• Working closely with the information council’s business liaisons to evangelize enterprise data governance within the organization

The success of an information governance organization depends on having the right candidate at the right level in the CDO role.

The data stewardship community function focuses on implementing the information governance policies and processes. It works with the end users to define the business and technical metadata, provides the data quality measures and ranges to be managed to (and performs data quality audits), and ensures that the end users are getting as much value as possible out of the data. The next section of this chapter covers those responsibilities for data stewards in greater detail.

The data stewardship community can be tightly aligned as a group or aggregated by organizational areas, as shown in Figure 1.3.
Where the data stewards are aligned to the organizational area, they are often “solid” lined (direct reporting) to that organizational area, and “dotted” line (indirect reporting) to the IGC.

**The Information Governance Reporting Models**

The success (or failure) of information governance initiatives are often a direct result of the alignment of the information governance organization within the enterprise. There are many reporting models, but there three are typically found:

- **Aligned to the chief financial officer (CFO)**—In this reporting model, the CDO and the IGC direct report to the CFO. This model has been used in both the manufacturing and life science industry. There are many benefits to this alignment that include tying budgets to adherence to information governance standards, tight alignment to financial management reporting (business metadata management), and the usage of financial information (data security).

- **Aligned to the chief risk officer (CRO)**—This model is most prevalent in the financial services industry, where adherence to government regulatory requirements and mandates is tightly tied to the common set of data definitions and the ability to demonstrate data lineage (e.g., Sarbanes-Oxley).
• **Aligned to the chief information officer (CIO)**—In this reporting model, the CDO and the IGC direct report to the CIO. One of the advantages of reporting to the CIO is the tight alignment to the development and maintenance of the data assets within the organization. Among the disadvantages of information governance organizations aligning within the IT is that business functions tend to view those organizations as technical only and discount the importance of the discipline. This leads to issues in the enforcement of (and lack there of) information governance standards and guidelines within the business functions.

**Data Stewardship Component**

Data stewardship is the “people” aspect of information governance that directly interfaces with the creators and users of data. Data stewards support, maintain, and execute the policies and procedures instituted by the IGC. Data stewards are often organized in communities that are aligned either (or both) by functional areas, such as customer or product, (and) or by departmental areas, such as accounting or marketing. Most information governance tasks discussed in this text are either directly performed by or influenced by data stewards.

**Typical Data Stewardship Responsibilities**

A data steward’s responsibilities vary widely from organization to organization based on the structure of the information governance process, the maturity of information governance within the enterprise (e.g., perceived important and authority granted to the information governance organization), and how the enterprise has organized its IT function. Typical data stewardship responsibilities are shown as follows and categorized by how data is created, organized, managed, and monitored. These responsibilities include the following:

• **Data stewardship creation responsibilities:**
  • Work with the business stakeholder and technologies in the business and technical definitions of data requirements
  • Ensure that the planned data has defined data quality criteria and ranges for critical data entities
  • Ensure that those definitions are captured and stored as metadata
  • Collaborate with IT data architects and modelers to ensure that the captured data requirements are structured correctly so that the intended business users gain the intended value
  • Collaborate with the business users and corporate security on data privacy requirements, user access control procedures, and data-retention policies
• **Data stewardship management responsibilities:**
  • Review and approve potential changes to the definitions and structures of the data, ensuring that those changes are appropriately maintained in the metadata management environment
  • Provide ongoing communications on the information governance organization, its policies, and processes
  • Assist/perform “road shows” on evangelizing information governance
  • Work with data management organizations on embedding information governance activities into ongoing processes and activities

• **Data stewardship monitoring responsibilities:**
  • Manage and communicate changes to data quality and security controls to business and technical stakeholders
  • Perform ongoing data quality and security audits on critical subject areas and application systems within the organization
  • Manage issues due to technical data quality and definitional understanding inconsistency, including data quality renovation projects

The breadth of information governance within the processes of an organization has led to the development of several types of data stewards and data stewardship-type roles. Most of these are segmented between business and technology roles, each with certain characteristics and responsibilities. The following sections provide a noncomprehensive list of the types of data stewards.

**Business Data Stewards**

Business data stewards focus more on the interaction of data with the executives and end users of a business function. They tend to focus on the data definition of base and aggregated data. For example, the business definition and calculation of return on net assets (RONA) can be a hotly contested definition between functional areas of an organization and a source of considerable time and effort for data stewards to develop common understandings and agreed to definitions to avoid perceived data quality issues and erroneous reporting. These business data stewardship roles include the following:

• **Departmentally focused data stewards**—These stewards tend to align into organizational areas such as accounting, finance, and marketing. They narrowly focus on the definition, creation, maintenance, and usage of data only within an organizational area. Often these data stewards are aligned closer to the executive of that organizational area than with the information governance organization (for example, finance data stewards that report directly to the CFO).
• **Functionally focused data stewards**—These stewards tend to align closer to the information governance organizations and are responsible for the definition, creation, maintenance, and usage of data for a functional area such as customer or product that may span many different organizations. For example, the customer domain may cross finance, accounting, marketing, production, and distribution. It requires an understanding of how the definition and process events that impact the concept of customer as a customer are processed from potential to purchaser of the organization’s goods and services. This broader organizational view almost always needs an information governance process to reconcile all the different organizational perspectives.

**Technical Data Stewards**

Technical data stewards focus more on the technical definition, creation, and maintenance of the data. They tend to report to IT, often the data management group, and provide the interface between IT and the business functional areas. These roles include the following:

• **Analytic data stewards**—These data stewards focus on the definition, maintenance, and usage of data generated from BI environments. Because much of this data has been transformed from its raw state through calculations and aggregations, one of the major tasks of these stewards is ensuring that the stakeholders agree to the common definitions and calculations of this data. They often work with the IT developers and end users in the definitions of the key performance measurements, calculations, and aggregations that make up the reporting. These are also the data stewards that work very closely to ensure that the information used for regulatory reporting meets the technical requirements of correctness and security.

• **Metadata management stewards**—These individuals have a very specific data stewardship focus on capture, maintenance, and versioning of the various types of business and technical metadata. They play a role that transcends IT’s data management organization and the IGC in managing the metadata environment. For those organizations that have established a commercial or homegrown metadata management repository, these data stewards are responsible for the capture, versioning, and maintenance of the different types of metadata. Later this chapter provides a broader definition of the different types of metadata that are created and managed.

• **Data quality analysts**—These specific-purpose data stewards concentrate on the data quality aspects of a functional or organization area within an information governance organization. They assist in the definition of the data by focusing on what the data quality criteria are for critical data elements (for example, what the technical and business domains and ranges are). They also approve the critical data elements to meet the project’s data quality requirements. They manage and perform the ongoing data quality audits and renovation projects on behalf of the information governance organizations.
Note that these are simply types of roles; in certain organizations, the same individual will perform any number of these data stewardship roles. The number and definition of the types of roles are also a function of the information governance maturity within an organization. The more mature the information governance, the more delineation will be found within the types of data stewardship roles.

**Common Characteristics of Data Stewards**

Regardless of type, certain common characteristics are found in all data stewards, such as a deep understanding of the underlying data and the processes and business rules that create that data; they are usually the data experts. Good data stewards tend to have deep industry expertise; they are very experienced practitioners in the industries that they work in. For example, a healthcare data steward understands the critical nature of ICD-10 codes, whereas a banking data steward is familiar with the regulatory requirements of the Dodd-Frank Act. They are by nature data evangelists, often with a deep passion for the data and its definition. Good data stewards tend to be 40% trained and 60% passion.

Understanding that the data steward is the performing “people” part of information governance ensures that when information governance activities and tasks are performed in development and ongoing operations, data stewards will in most instances be a primary or secondary performer.

**Data Quality Management Component**

Data quality management is the definition, supervision, and when necessary, renovation of data to the business and technical ranges. Data quality management is one of the most visceral aspects of information governance. It is also “threads” through each of the “people, process, and technology” aspects of information governance. For example, organizational reactions to perceived or real data quality issues have cost organizations millions of dollars in regulatory fines, cost executives their positions, and are one of the primary reasons companies start information governance initiatives. However, despite all the press, it is still one of the least understood areas of information governance.

**What Is Data Quality?**

Data quality is the commonly understood business and technical definitions of data within defined ranges. It is measured by how effectively the data supports the transactions and decisions needed to meet an organization’s strategic goals and objectives, as embodied in its ability to manage its assets and conduct its core operations.

The level of data quality required to effectively support operations will vary by information system or business unit, depending on the information needs to conduct that business unit’s operations. For example, financial systems require a high degree of quality data because of the importance and usage of the data, but a marketing system may have the latitude to operate with a lower level of data quality without significantly impacting the use of the information in
measuring marketing success. Because the purpose varies, so does the bar used to measure fitness to purpose.

**Causes of Poor Data Quality**

Causes for bad data quality can be categorized as business-process and technology-process data quality issues, as demonstrated in Figure 1.4.

<table>
<thead>
<tr>
<th>Cust No</th>
<th>Customer Name</th>
<th>Product</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Ms. John Smith</td>
<td>Seats</td>
<td>$1,200</td>
</tr>
<tr>
<td></td>
<td>Sam Reilly</td>
<td>Chairs</td>
<td>$2,300</td>
</tr>
<tr>
<td>11</td>
<td>Jack Jones</td>
<td>Stools</td>
<td>$1,750</td>
</tr>
<tr>
<td>13</td>
<td>Charles Nelson</td>
<td>Tables</td>
<td>$A,AAA</td>
</tr>
</tbody>
</table>

**Figure 1.4** Examples of bad data quality types

Technology-driven poor data qualities are those types that are caused by not applying technology constraints on either the database or data integration. These types include the following:

- **Invalid data**—Data that is incorrect in that field. For example, by not applying constraints, alphanumeric data is allowed in a numeric data field (or column).

- **Missing data**—Data that is missing in that field. For example, by not applying key constraints in the database, a not-null field has been left null.

Business-driven bad data qualities are those types that are caused by end users inaccurately creating or defining data. Examples include the following:

- **Inaccurate data**—Invalid data due to incorrect input by business users. For example, by inaccurately creating a record for Ms. Anthony Jones, rather than for Mr. Anthony Jones, poor data quality is created. Inaccurate data is also demonstrated by the “duplicate data” phenomenon. For example, an organization has a customer record for both Anthony Jones and Tony Jones, both the same person.

- **Inconsistent definitions**—Inconsistent data is where stakeholders have different definitions of the data. By having disparate views on what the definition of poor data quality is, perceived bad quality is created. For example, when the sales department has a different definition of customer profitability than the accounting department.

**The Data Quality Framework**

Most EIM functions have an architecture or framework by which to understand that function; data quality is no exception. The data quality framework illustrated in Figure 1.5 is a multidimensional
The Information Governance Component Framework

reference model to explain and define data different dimensions of data quality. The first dimension defines the key data quality elements, or what data within an organization or application is important to measure quality. The business and technical dimensions provide the rules that measure how well a data element meets a company’s data quality goals and ultimately provides trusted and critical information.

![Diagram of data quality dimensions]

**Figure 1.5** The dimensions of the data quality framework

Understanding all four aspects of this framework will help you determine what information governance activities and tasks must be performed to ensure the levels of data quality desired by an organization.

**Key Data Quality Element Dimension**

Within an organization, certain data elements are critical to the business and so the data quality of such should be identified, defined, and measured. These key data elements can be both base element data (for example, customer name) as well as derived data (for example, net profit).

These key data quality elements are often defined as such during data definition activities such as data modeling. Once identified as a key data quality element, the technical and business data quality criteria for that element are identified and defined in terms of ranges of compliance to requirements of a business. For instance, the key data quality element birth date has a business data quality criteria defined as a date range, as follows:

Birth date = Range: from 0 to 140
This business user-defined range reflects the probability that most people simply do not live beyond 140 years.

Although a relationship exists between relational key constraints, mandatory data, and key data quality elements, that relationship is not one to one. Not all mandatory and constraint data is necessarily key data quality data.

For instance, a customer ID column may be both mandatory and a primary key constraint, but not a key data quality element based on that element’s importance to the organization.

**Business-Process Data Quality Dimension**

The business-process data quality dimension refers to the data quality criteria based on the business definition and business rules defined within the data. It contains the business defined ranges and domains that are a direct result of a business decision.

It is the lack of formal definition or misunderstanding of the different interpretations that create the inconsistent definitions and different business rules for similar data within each line of business (LOB), with each LOB having its own understanding of what that data element is. For example:

- Marketing definition of net assets = Assets – Expenses
- Finance definition of net assets = Assets – Expenses + Owners equity

Hence, with disparate views on what the definition and business rules of a data quality element are, when information is compared from different LOBs, the perception of bad quality is created, as shown in Table 1.1.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Examples of Poor Business Data Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitional</td>
<td>The data element has a commonly agreed-upon enterprise business definition and calculations.</td>
<td>Return on net assets (RONA), net present value (NPV), and earnings before interest, taxes and amortization of goodwill (EBITA) are calculated using different algorithms/equations and using different source data for each algorithm/equation for multiple departments within an enterprise.</td>
</tr>
</tbody>
</table>

Applying a consistently agreed-upon *common* business definition and rules against the data elements provides the insurance against inconsistent data quality issues.

It is the management of the common understanding of business definitions throughout the data stewardship community that is so critically important to not have misunderstood reporting issues.
**Technical-Process Data Quality Dimension**

The technical-process data quality dimension refers to the data quality criteria found in the technical definition of the data (for example, as defined in both the entity integrity and referential integrity relational rules found in logical data modeling). Table 1.2 describes key aspects of this dimension.

**Table 1.2  Technical Dimensions of Data Quality**

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Examples of Poor Technical Data Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>The data element passes all edits for acceptability.</td>
<td>A customer record has a name that contains numbers.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The Social Security Number field should be a numeric integer but is populated with alphanumeric characters instead.</td>
</tr>
<tr>
<td>Unique</td>
<td>The data element is unique; there are no duplicate values.</td>
<td>Two customer records have the same Social Security number.</td>
</tr>
<tr>
<td>Complete</td>
<td>The data element is always required or required based on the condition of another data element.</td>
<td>A product record is missing a value such as weight.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Married (y/n) field should have a non-null value of y or n, but is populated with a null value instead.</td>
</tr>
<tr>
<td>Consistent</td>
<td>The data element is free from variation and contradiction based on the condition of another data element.</td>
<td>A customer order record has a ship date preceding its order date.</td>
</tr>
<tr>
<td>Timely</td>
<td>The data element represents the most current information resulting from the output of a business event.</td>
<td>A customer record references an address that is no longer valid.</td>
</tr>
<tr>
<td>Accurate</td>
<td>The data element values are properly assigned (e.g., domain ranges).</td>
<td>A customer record has an inaccurate or invalid hierarchy.</td>
</tr>
<tr>
<td>Precise</td>
<td>The data element is used only for its intended purpose, i.e., the degree to which the data characteristics are well understood and correctly utilized.</td>
<td>Product codes are used for different product types between different records.</td>
</tr>
</tbody>
</table>

Each of these technical data quality rules are enforced against the key data quality elements with different methods. Many of the rules are enforced with simple relational database rules such as entity and referential integrity. For instance, the precise dimension is enforced in the relational database by applying the primary key constraint.

Within each of these dimensions, technical data quality rules are applied against key data quality elements, as shown in Figure 1.6.
Table: Data Quality Criteria Work Book

<table>
<thead>
<tr>
<th>Customer Column Name</th>
<th>Valid</th>
<th>Unique</th>
<th>Complete</th>
<th>Consistent</th>
<th>Timely</th>
<th>Accurate</th>
<th>Precise</th>
<th>Enterprise Business Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cust_Id</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>The unique identifier assigned to a Customer.</td>
</tr>
<tr>
<td>Cust_First_Name</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Specifies the first name of the Party.</td>
</tr>
<tr>
<td>Cust_Last_Name</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Specifies the last name of the Party.</td>
</tr>
<tr>
<td>Gender</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gender of the customer.</td>
</tr>
</tbody>
</table>

Data Quality Checkpoints

Capturing and renovating bad data that has been defined in the context of the data quality framework can be prevented by determining key data quality criteria and building those rules into data quality checkpoints. There are two types of data quality checkpoints:

- **Technical data quality checkpoints**—Technical data quality checkpoints define the data quality criteria often found in both the entity integrity and referential integrity relational rules found in logical data modeling. They address the invalid and missing data quality anomalies. Technical data quality criteria are usually defined by IT and
information management subject matter experts (SMEs). An example includes the primary key null data quality checkpoint.

- **Business data quality checkpoints**—The business data quality checkpoints confirm the understanding of the key data quality elements in terms of what the business definition and ranges for a data quality element are and what business rules are associated with that element. Business data quality checkpoints address the inaccurate and inconsistent data quality anomalies. The classic example of a business data quality check is gender. A potential list of valid ranges for gender is Male, Female, or Unknown. This is a business definition, not an IT definition; the range is defined by the business. Although many organizations find the three values for gender sufficient, the U.S. Postal Service has seven types of gender, so their business definition is broader than others.

### Types of Data Quality Processes

The final aspect of the data quality framework are those processes that ensure good data quality or prevent bad quality from being created and those that find bad data quality for renovation.

Ensuring data quality is typically a result of solid adherence to the definition of data quality criteria from both a business process and data design perspective. As a result, there are preventive data quality best practices that focus on the development of new data sources and integration processes, and there are detective data quality best practices that focus on identification and remediation of poor data quality. Both of these types are found in the tasks and steps of the data quality life cycle, which is discussed in Chapter 11, “Ongoing Data Quality Management Processes.”

The understanding of what data quality is, the framework for which it is defined, and how to capture data quality is critical to understanding one of the important “process” components of information governance, especially in terms of ensuring the right data quality processes are built and then monitored in ongoing operations.

### Metadata Management Component

The metadata management component is one of the process and technology aspects of information governance that captures, versions, and uses metadata to understand organization data. It is the “database” for data stewards and other types of users to store, maintain, and use the business and technical definitions of the organization’s data.

What is metadata? Metadata is defined as “data about data,” but it can also be explained as another layer of information created to help people use raw data as information. Metadata provides context to raw data; it is the business and technical rules that provide that particular data element meaning, as illustrated in Figure 1.7.
What is Metadata?

**Data Element Name:** Customer Profitability

**Business Definition:**
It is a key reporting performance measure that calculates the profitability of the organization's customers.

**Technical Definition:**
- **Data Type:** Real
- **Length:** 10.2
- **Source or Calculated:** Calculated
- **Calculation:** Total Customer Revenue - Expenses

**Figure 1.7** Types of metadata: Business and structural

Metadata is created whenever data is created, either in transaction processing, master data management (MDM) consolidation, or BI aggregations. Each event creates a type of metadata that often needs to be captured and managed. For example, when a data element is created, it contains information about what process was used to create it, along with rules, formulas, and settings, regardless of whether it is documented. The goal is to capture this metadata information at creation to avoid having to rediscover it later or attempt to interpret it later.

The discipline of metadata management is to capture, control, and version metadata to provide users such as data stewards the ability to manage the organization’s data definitions and data processing rules in a central location. The application to capture, store, and manage metadata is a metadata repository, which is a metadata “database” for use by stakeholders such as data stewards.

Metadata can be composed of any information that describes the actual data itself. For data warehousing purposes, metadata has been classified based on the purpose created and the functions it is used for and can be classified into the types or categories. In each of these categories, there are relationships. For example, navigational, structural, and analytic all require the business definitions in the business metadata to provide context to the data, as demonstrated in Figure 1.8.
Figure 1.8  The categories of metadata

**Business Metadata**

The business category of metadata defines the information that the data provides in a business context. Examples of business metadata include subject area definitions (e.g., product), entity concept definitions, business attribute names, business attribute definitions, business attribute valid values, data quality rules, and business rules. Business metadata is found in transactional data master data. One of the primary sources of business metadata includes conceptual data models, logical data models, and business process rules engines.

**Transactional Metadata**

Transactional metadata contains the business and technical data definitions and business rules used in creating transactional systems. Transactional metadata is the source of all downstream uses of information, and when it is poorly defined or enforced, it is the major source of data quality issues.
Structural Metadata

Structural metadata contains the logical and technical descriptions of the permanent data structures within the EIM infrastructure. This metadata includes structures such as flat files and hierarchical and relational databases. Structural metadata contains both logical and technical metadata, as shown in Figure 1.9.

Logical metadata consists of data models and entity, attribute, and relationship metadata. A level of overlap exists between business and logical metadata (for example, business attributes and physical attributes). Business attributes are defined by the business to describe an aspect of
an entity. A physical attribute is defined by a data modeler or application database administrator to describe an aspect of the physical store of data. Some organizations only retain and manage the one type.

The technical metadata is the physical structures themselves (for example, databases/file groups, tables/views/files, keys, indices, columns/fields, source columns/fields, and target columns/fields). Often this type of information is found in Database Definition Language (DDL).

**Navigational Metadata**

Navigational metadata describes the process rules and data formats of the data extraction, transformation, and movements, as illustrated in Figure 1.10. Examples of navigational technical metadata are derived fields, business hierarchies, source columns and fields, transformations, data quality checkpoints, target columns and fields, and source and target locations. Primary sources of navigational metadata include data profiling results, data mappings, logical/physical data integration models, and data quality criteria workbooks.
Commercial data integration software vendors have addressed navigational metadata from two perspectives:

- **Integrated software suites**—IBM, Ab Initio, and Informatica have integrated profiling and data analysis tools into their design and development suites. This includes data mapping.
- **Metadata repositories**—The same vendors have metadata repositories for navigational metadata as well as the capabilities to integrate other types, which is discussed later in the chapter.

**Analytic Metadata**

Analytic metadata, shown in Figure 1.11, consists of the metadata that is used in a reporting and ad hoc environment and includes the following:

- **Report data elements**—Within the report itself, the definition of the report-level data elements displayed on the report or in the ad hoc query environment is metadata to be created and managed. These elements are often the same technical and business definitions as the data warehouse or dimensional data mart.

**NOTE**

However, these data elements often have changed technical and business metadata that is different from the data warehouse environment, leveraging the ability of the commercial analytic tool metadata capabilities. These changes should be captured and documented from both a data stewardship and metadata management perspective.

- **Report-level aggregations and calculations**—Most commercial analytic tools provide the ability to build aggregations and calculations at the report level.
- **Report layout and report navigation metadata**—This technical metadata describes the layout of the report, the fonts to be used, and how the data should be shown and navigated.
Primary sources of analytic metadata include OLAP and reporting packages metadata environments.

**Master Data Metadata**

Master data metadata crosses both transaction and analytic application definitions that describe the core business domains of an organization. Master data provides transaction and analytic data the context of the organization for core domains such as party-customer, product, and account, as shown in Figure 1.12.
Operational Metadata

The operational category of metadata describes the transaction and data integration application’s job description through statistics, giving a full technical view of the environment. Examples of operational metadata include jobs statistics and data quality check results.

Whereas the prior categories are primarily used by business users, data stewards, and data management professionals, operational metadata is used by production support and systems administration for troubleshooting and performance tuning.

Sources of operational metadata include transaction and data integration job logs being generated either by the data integration jobs or the production scheduler.

Metadata Users

Metadata provides value at a variety of levels to a range of users but can typically be divided into three categories:

- **Business users**—Business users of metadata need to understand the business meaning of the data in the systems they use. In addition, they need to know the business rules and data access rules that apply to the data. Data stewards (either business or technology) are usually classified as business users due to the creation, maintenance, and usage patterns of metadata.

- **Technology users**—IT professionals who are responsible for planning and building the transaction and analytic systems need to understand the end-to-end picture of the data to manage change. These users leverage the technical metadata for the technical information about the data environment, such as physical data structures, extract-transform-load
rules, reporting information, and impact analysis. Examples of technology users include data modelers, service-oriented architecture (SOA) architects, data-integration architects, BI architects, designers, and developers.

- **Operational users**—IT operational professionals are those who are responsible for day-to-day operation of the data environment and are users of operational metadata. Operational metadata can assist them in identifying and resolving problems as well as managing change in the production environment by providing data information about the data integration processing and job processing impact analysis.

**Managing Metadata**

Because metadata is created in many places during the development of a system, it is important to understand and govern all the categories of metadata in the metadata life cycle. Information management professionals have had the goal of a centrally managed metadata repository that governs all metadata, but that vision is difficult to achieve for a variety of factors. The reality is that metadata is created in many different tools used to develop data structures and process that data, as shown in Figure 1.13.

![Figure 1.13 Centrally managing sources of metadata](image)
At best, a centralized metadata repository should enhance metadata found in local repositories. A metadata repository should consider the following:

- **Where it will be stored**—Identify the data store requirements (e.g., commercial metadata repository, homegrown relational database).
- **What will be stored**—Identify metadata sources.
- **How it will be captured**—Identify load mechanism, CRUD (create, read, update, delete) requirements, administration requirements, and audit and retention requirements.
- **Who will capture the data**—Identify the roles and responsibilities for managing the repository and levels of users.
- **When it will be captured**—Identify capture frequency, history, and versioning considerations.
- **Why it will be captured**—Identify the benefits of the requirements and the specific questions this metadata will answer and provide reporting/browsing requirements.

Metadata is an organization’s “encyclopedia” of business and technical definitions for use by data stewards and other key users. Capturing and updating metadata is a very visible and important set of activities in performing information governance.

Understanding that the data steward is performing the “people” part of information governance ensures that when information governance activities and tasks are performed in development and ongoing operations, data stewards will in most instances be a primary or secondary performer.

**Privacy and Security Component**

The privacy and security component covers all three of the people, process, and technology aspects of information governance to address who has create, read, update, and delete privileges of organizational data. There have been security requirements for data since the beginning of IT, with access and file security on mainframes with ACF2 and RACF security packages. This was further refined with the advent of relational database technologies with role- and column-level security and “locking data” down with schema-level security roles.

Privacy has taken on an equal if not more important (from a legal liability perspective) role with the integration of organizations’ intranets with the external Internet. The ability for nonstakeholders to access critical financial, customer, and employee data has spawned legislation such as personally identifiable information (PII) laws on how data can and cannot be used to identify, contact, or locate an individual. Another example is in the healthcare industry in the Health Insurance Portability and Accountability Act of 1996 (HIPAA) privacy and security law, which seeks to ensure the privacy and security rights of an individual’s health information. These and other such laws have made the role of information governance even more prominent.
A Broader Overview of Security

Information governance security “interlocks” with the broader IT security and general security functions at the data definition and usage level, as shown in the classic Venn diagram in Figure 1.14.

![Information Governance and Security Venn Diagram](image)

**Figure 1.14** Information governance security and privacy in the context of a broader security function

As with other information governance components, there is a framework that best describes how security and privacy “threads” into EIM functions, as shown in Figure 1.15.

<table>
<thead>
<tr>
<th>EIM Functions</th>
<th>Transactional Processes</th>
<th>Master Data Management</th>
<th>Business Intelligence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data</td>
<td>Process</td>
<td>Objects</td>
</tr>
<tr>
<td>Business Privacy &amp; Security Requirements</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Technology Privacy &amp; Security Requirements</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>External Requirements</td>
<td>PII</td>
<td>HIPAA</td>
<td>Other Regulatory Agencies</td>
</tr>
</tbody>
</table>

**Figure 1.15** Security and privacy framework
Each EIM functional component of the framework in Figure 1.15 requires a thoughtful analysis and implementation approach for each of the dimensions for the business, technical, and external requirements for privacy and security. For example, a healthcare organization’s member data that is collected through a website needs to consider the following:

- **Business privacy and security requirements**—Who with the healthcare organization is allowed to access that data?
- **MDM technical requirements**—What are the business, technical, and HIPAA (regulatory) rules for integrating this member data with other stores of member data?
- **Privacy and security requirements in analytic analysis**—How can the member data collected from the Web be used for member profiling without violating HIPAA?
- **Technical privacy and security requirements for the data warehouse**—What technical solution, such as database security, schema security, and user roles, will meet HIPAA requirements for healthcare member data?

Each EIM “functional layer” of data should be determined through stewardship processes in conjunction with the chief information security officer.

**Chief Information Security Officer**

The critical nature of security and privacy has placed the chief information security officer (CISO) in the IGC as a board member, as shown in Figure 1.3. The CISO works with the CDO in setting security and privacy policies and often works directly with data stewards on project and operational issues surrounding security and privacy. For example, a data steward may need to review proposed security standards with the CISO to ensure that they meet HIPAA requirements.

Understanding how privacy and security is defined for data based on the business, technical, and regulatory requirements is critical in performing information governance.

**Information Life Cycle Management Component**

Information life cycle management (ILM) covers the process and technology aspect of information governance that addresses the entire life cycle of a set of data, including creation, retention, and deletion. It covers the business rules on how long data is to be kept and in what format. Due to the very technical nature of ILM, it is as much a data management discipline as it is a component of information governance. Despite the commoditization of computing CPU and disk storage, retaining vast amounts of data that can easily be into the hundreds of petabytes can run in the range of $50 million to $100 million per year. Based on usage and legal requirements, data can be cycled from traditional “hot” storage to cheaper archived storage that can still be accessed as needed (thus saving considerable amounts of money).

It is important for data stewards to consider both the usage and the legal requirements in determining whether to archive or delete old data. For example, a telecommunications company’s
data warehouse is required to store 4 years of billing data in its data warehouse. However, for tax compliance, it is required for 7 years; so, a potential life cycle management plan for billing data would be 4 years online and then 3 years offline/archived. After 7 years, the data could be deleted. So, in most cases, the following formula can be used:

Data must be retained for whichever is greater: organizational retention requirements or regulatory retention requirements.

This area of information governance has become much more focused since it provides a much more manageable and cost-effective approach to storing vast amounts of data.

Information life cycle management is one more dimension to consider when defining data and performing data stewardship audits.

**Information Governance in the Context of EIM**

With each of the information governance components defined, the next step is to understand the relationship of information governance in the context of EIM in order to document how to perform these information governance activities and tasks in both project and ongoing operations.

Observe how the information governance components thread into each of the different EIM functions, as documented in Figure 1.16.
A fully functional information governance organization would be deeply embedded into the development and maintenance cycle of each of the EIM functions in order to determine, define, and audit the information governance aspects of the information environment.

The goal will be to ensure that each of the information governance component requirements are integrated into the EIM functions’ common design and development themes.

Summary
This chapter provided a review of each of the different components of information governance, as follows:

- **Information governance organizational component**—The information governance organizational component is the “people” aspect of the discipline that sets the policies for information governance and maintains a staff to support those policies in managing the development and ongoing usage of corporate information.

- **Data stewardship component**—Data stewards support, maintain, and execute the policies and procedures instituted by the IGC.

- **Data quality management component**—Data quality management is the definition, supervision, and when necessary, renovation of data to the business and technical ranges.

- **Metadata management component**—The metadata management component is one of the process and technology aspects of information governance that captures, versions, and uses metadata to understand organization data.

- **Privacy and security component**—The privacy and security component covers all three of the people, process, and technology aspects of information governance to address who has create, read, update, and delete privileges of organizational data.

- **Information life cycle management component**—Information life cycle management covers the process and technology aspect of information governance that addresses the entire life cycle of a set of data, from creation, retention, and deletion.

Each of these components is threaded in each of the EIM functions, transactional processing, MDM, and BI.

The policies and requirements for each of the components must be instantiated to achieve the context and quality of the data needed for transaction and analytic processing within the organization.

This chapter defined each of the information governance components that need to be performed to achieve the context and quality of the data needed for transaction and analytic processing within the organization. Chapter 2, “Other Core EIM Functions,” reviews, at a high-level, the transaction, operational, and analytic functions of EIM so that you can understand where the requirements for the information governance components will interface.
End-of-Chapter Review Questions

1. What is the formal definition of information governance?
2. Fill in the blank: The information governance organizational component is the ____ aspect of the discipline that sets the policies for information governance and maintains a staff to support those policies in managing the development and ongoing usage of corporate information.
3. What are some of the reasons why organizations fail in their information governance efforts?
4. What types of data stewards are discussed in this chapter?
5. What are the four dimensions of the data quality framework?
6. What are the four causes of poor data quality?
7. When is metadata created?
8. What are the three types of privacy and security requirements that need to be considered when determining the creation and usage of data?
9. What is the definition of information life cycle management?
10. True or false: Information governance security “interlocks” with the broader IT security and general security functions.

Essay Question
Explain what information governance components are in place in your organization today.

For Further Reading

General Data Management

Enterprise Information Management
http://www.amazon.com/Practical-Implementing-Enterprise-Information-Management/dp/1934938920/ref=sr_1_sc_1?s=books&ie=UTF8&qid=1354892060&sr=1-1-spell&keywords=stephen+t+boschulte
http://www.amazon.com/The-Enterprise-Information-Architecture-Systems-Based/dp/0137035713/ref=sr_1_sc_1?s=books&ie=UTF8&qid=1354892060&sr=1-1-spell&keywords=the+art+of+enterprise+information+architecture

Master Data Management
http://www.amazon.com/MASTER-DATA-MANAGEMENT-GOVERNANCE/dp/0071744584/ref=sr_1_1?ie=UTF8&qid=1354892176&sr=1-1&keywords=mdm
Enterprise Architecture
http://www.amazon.com/Enterprise-Architecture-Planning-Developing-Applications/dp/0471599859/ref=sr_1_4?s=books&ie=UTF8&qid=1354892283&sr=1-4&keywords=enterprise+architecture

http://www.amazon.com/Patterns-Enterprise-Application-Architecture-Martin/dp/0321127420/ref=sr_1_3?s=books&ie=UTF8&qid=1354892283&sr=1-3&keywords=enterprise+architecture
This page intentionally left blank
Index

A
accounts, MDM requirements, 178
accurate technical data quality dimension, 21, 432
ad hoc queries, 53
administration
data ownership, 412-413
metadata repository, 489-491
administrator role, 490-491
configuration management, 490
technical/functional tasks, 489-490
agile method, 102
BI projects, applying, 216-217
custom transactional. See custom transactional processing projects
information governance, threading, 103
iterative phases, 103
analytics
BI, 227
architecture, 53-54
business impact, 233-234
data integration, 230-233
data warehouse requirements, 228-230
logical design, 234-235
physical design, 244-245
requirements, 228
SDLC, 213-214
transformations, 216
unit testing, 245
big data reviewing, 356-357
use cases, 360
change management, 514-515
claims cost solution, 262
customer, 341-343
data stewards, 16, 80
data structures
atomic data warehouses, 55
data marts, 55
development tasks, 253
gap analysis
BI architecture, 224-225
COTS transactional processing applications, 132
MDM architecture asset, 174-175
organizational skills, 173, 226
healthcare company BI project
logical data integration, 266-269
logical data warehouse design, 263-265
logical design, 262-263
physical data integration, 269
providers cost, 263
requirements, 260-261
healthcare package metadata, 500
implementations, 50-51
layer testing, 241
MDM, 50-51
change management impact, 186
data integration requirements, 182-184
functionality prototypes, 184-185
implementation, 177
nonfunctional requirements, 185
requirements. See MDM, requirements
information governance considerations, 52
logical design, 234
analytics, 234-235
data integration, 238-240
data warehouses, 236-238
metadata management, 212
organizational integration, 387-388
sales and marketing processes, 52
SDLC tasks and activities, listing of, 587
strategies, 220
architectural, 223-225
client business/IT environment assessment, 220-221
ongoing organizational program, 221-223
technical design, 243
analytics, 244-245
data integration, 249-252
data warehousing, 246-249
training, 252
testing, 241-243
types, 241-242
user acceptance, 243
transformations, 215-216
big data
analytics, reviewing, 356-357
clinical trials project, 362-363
control phase, 363
new clinical research data definitions, 364-365
quality control, 365
usage phase, 366
control, 334, 350
captures, 350-352
data definition tasks, 352-354
defined, 619
quality, 354-355
data ownership
overview, 333
rules, applying, 334

data stewards, 332-333
dimensions, 60, 330
discovery, 330-331, 345
defined, 619
new sources, analyzing, 345-346
quality control checkpoint, 349-350
security, 348-349
subject areas, 346-348
information life cycle management, 61
integrated, 330, 621
latency, 339
life cycle, 343-344
new patterns, 340
analytic enhancements, 341-343
real-time operational systems integration, 340-341
opportunities, 339
immediacy, 339-340
real-time decisions, 340
reviewing, 355-357
overview, 60
ownership, 61, 352
pharmaceutical company clinical research project
WBS, 617-616
quality, 61
security/privacy, 61
semantic federation, 339
SLC tasks and activities, listing of, 616
source access, 356
usage patterns, 334-335
integration, 337
real-time operations, performing, 338
staging for discovery, 335-336
usage phase, 355
defined, 619
new integrated data, identifying, 357-359
opportunities, reviewing, 355-357

prototyping, 361-362
security, 360
use cases, developing, 359-360
blueprints. See architecture build phase. See development business
continuity, 207
BI SDLC, 232
MDM, 184
data quality checkpoints, 23
data stewards, 15-16
intelligence. See BI metadata, 25, 473
capturing, 500-501
logical data warehouses, 237-238
support, 486
users, 30, 478
modeling
agile transaction processing SDLCs, 110-111
business process models, defining, 110
defining, 110, 408
requirements, 111-112
rules, capturing, 110-111
processes
baseline, reviewing, 131
baseline current, 131
data quality dimension, 20, 433-434
developing/leveraging, 133
high-profile/high-risk, prototyping, 135
rules, capturing, 408
services, MDM requirements, 181
subject areas, 619
buying metadata repositories, 491

C
call center management, 341
capabilities maturity model (MDM), 169
capturing metadata
business/technical, 500-501
data stewards, 418-419
healthcare claims example, 499-500
navigational, 501
overview, 481-482
repository, 483
case studies
healthcare company, 623-624
pharmaceutical company, 624-625
CASE tools, 478
CDOs (chief data officers)
EDM
integration, 383-384
operational reporting, 385-386
IGC
leaders ongoing activities, 373-374
stakeholder operational activities, 375
information governance role, 11-12
operational responsibilities
data quality, 377-379
data stewardship, 376-377
metadata management, 379-381
security, 381-382
roles, defining, 71-72
span of control, 389
CFOs (chief financial officers), 13
change management
analysis, 514-515
challenge recommendations
enterprise, 510
operations, 512
project level, 513
challenges
executive, 508
operations, 509
projects, 509
transition, 507
COTS transactional applications, 139-141
development team
performance objectives, 140
end-user security roles, 141
information governance organization requirements, 141
operations/maintenance teams, 140
organizational requirements, 139
stakeholder management, 139
custom transactional applications, 123-124
deployment, 517
design, 515-516
development, 516
importance, 505-507
MDM impact assessment, 186
change requests, managing, 123-124
checkpoints (data quality), 22-23
chief data officers. See CDOs
chief financial officers (CFOs), 13
chief risk officers (CROs), 13
CIOs (chief information officers), 14
CISO (chief information security officer), 34
responsibilities, 78-79
security and privacy role, 34
security reporting, 381-382
claims
costs
analytics solution
(healthcare project), 262
logical dimensional data model, 264
reengineering project, 124
challenges, 124
initiate phase plan, 126-128
planning, 125-126
WBS, 554
class diagrams, designing, 115, 408
cleansing
master data, 136-137
transactional data, 135-136
clean staging landing zones, 58
clinical trials project, 362-363
control, 363
new clinical research data definitions, 364-365
quality, 365
usage phase, 366
cloud storage, 277
coexistence architectural style, 49
common vocabulary task
capturing, 105-106
updating, 111
complete technical data quality dimension, 21, 432
compliance
audit RASIC chart, 93
external reviews, 375
healthcare data quality
ICD-10 compliance project, 461
H00-H59 criteria, confirming, 463
ICD-10 H00-H59 code extensions review, 463
ICD-10 master data, 461
remediation, 464-466
scope, 462-463
HIPAA project, 397-398
ILM
defining, 281-282
legacy claims renovation project, 315-317
monitoring, 299-300
renovation projects, 304
validating, 283-284
internal reviews, 375
reporting, 374
conceptual data models
data steward task, 405
defining, 106
integration, 410
configuration items, changing/delivering, 123
consistency
  measures, 447
  technical data quality
dimension, 21, 432
consolidation architectural style, 49
contract requirements, 178
control
  big data, 350
captures, 350-352
data definition tasks,
  352-354
defined, 619
  quality, 354-355
clinical trials project, 363
  new clinical research data
definitions, 364-365
  quality, 365
conversion functional
specifications, 134
core business model processes
finance and operations
  BI, 52
transaction processing,
  42-43
orders and production, 45-46
sales and marketing
  BI, 52
transaction processing,
  43-45
transaction processing, 42
cost assessments, 304
COTS transactional processing
projects, 129
architecture, 130
  baseline business
  processes, reviewing,
  131
  baseline current process
  activity, 131
blueprinting completion,
  132-137
blueprint walkthrough,
  138
gap analysis, 132
change management, 139-141
development team
  performance objectives,
  140
end-user security roles, 141
information governance
  organization
  requirements, 141
operations/maintenance
teams, 140
organizational
  requirements, 139
  stakeholder management,
  139
finance consolidation project,
  149-150
blueprint breakdown,
  152-157
planning, 150-151
mapping and construction
phase, 144-148
  components testing,
  147-148
core technical
  specifications/
development/tech and
  functional unit test, 146
development, 148
forms, 146
information life cycle
  requirements, 147
master data policies, 145
process configuration/
  confirmation, 144-147
reports, 146-147
stakeholder readiness
  survey, 144
organizational
  responsibilities, 390
pharmaceutical company
  finance consolidation
  project, 564
processes, buying, 129-130
SDLC activities and tasks,
  listing of, 561
testing, 142-143
  project metadata,
  capturing, 143
  security test cases,
  142-143
strategies, 142
CROs (chief risk officers), 13
CRUD (Create, Read, Update,
  and Delete) services, 189
defined, 621
MDM
  design, 189
development, 197
pharmaceutical customer
  project, 206
customers
  analytics, 341-343
  lifetime value, 342-343
  segmentation, 342
custom transactional processing
projects
  agile SDLC activities and
  tasks, listing of, 543
  claims reengineering project,
  124
  challenges, 124
  initiate phase plan,
  126-128
  planning, 125-126
COTS
  pharmaceutical company
  finance consolidation
  project WBS, 564
  SDLC tasks and activities,
  listing of, 561
data steward tasks, 407-408
development, 109
  architecture, 112-114
  business modeling,
  110-111
  construction phase, 120
design, 115-119
development, 120-121
elaboration phase,
  114-115
final approval signoff, 114
inception phase, 109
requirements, 111-112
testing, 119-122
healthcare claims
  reengineering project WBS,
  554
  initiate phase, 105
  construction iterations
  planning, 108
organizational requirements, 107-108
requirements, 105-106
solution architecture, 106-107
transition iteration, 122
configuration and change management tasks, 123-124
release and deployment tasks, 122
testing, 122

D
dashboards, 54, 448-450
data
big. See big data integration
BI architecture, 57-59
BI requirements, 230-233
conceptual model, 410
data quality life cycle, 443-445
development, 254
logical, 238-240
logical design, 266-269
MDM requirements, 182-184
physical, 269
physical design, 249-252
transformations, 215
lineage, 619
loads, 622
healthcare company BI project, 268
MDM, 165, 192-194
one-time initial logical design, 240
processes, 59
publish landing zones, 59
marts, 55
master. See MDM
migration, 118
stores, 621
landscape review, 286-287
master, 621
MDM, 190-192, 197-198
problems, reviewing, 303
renovation processes, developing, 305-306
warehouses
BI architecture, 55-57
BI requirements, 228-230
development, 253
healthcare company BI project, 263-265
healthcare data quality criteria, 500
layer testing, 241-242
logical design, 236-238
physical design, 246-249
transformations, 215
databases
custom transactional applications, creating, 116-117
development anomalies, 439
constraints, 440
data quality-specific tasks, 441-443
define phase
data quality life cycle, 436
criteria, 436-437
data integration, 443-445
development thread tasks, 438-439
modeling and database design, 439-443
preventive processes, 437-438
scope, 436
ILM, 284
data retention criteria, 285-286
data retention technology review, 288-289
data store landscape review, 286-287
implementation plan development, 289-291
definitional business data quality dimension, 433
deletion reports, 308
departmental data stewards, 15
dependency diagrams, 133
deployment
BI, 255
final preparations, 256-257
testing cycles testing, 255-256
change management, 517
COTS transactional applications, 148
data quality remediation, 460-461
MDM, 200-201, 498-499
policies, 395-396
design
BI logical, 234
analytics, 234-235
data integration, 238-240
data warehouses, 236-238
testing, 241-243
BI technical, 243
analytics, 244-245
data integration, 249-252
data warehousing, 246-249
training, 252
change management, 515-516
class diagrams, 408
custom transactional applications, 115-119
class diagram, 115
databases, generating, 116-117
data migration, 118
metadata, capturing, 118-119
physical data models, creating, 116
retention requirements, 117
service components, 118
user interfaces, prototyping, 115-116
databases, 439-443
data quality remediation, 457-459
healthcare company BI project, 269
logical BI
    analytics, 262-263
data integration, 266-269
data warehouses, 263-265
master data cleansing, 136-137
MDM
    CRUD components, 189
data load components, 192-194
data store model, 190-192
hub user interfaces, 187-188
orchestration components, 188
package configuration, 195
prototypes, 186-187
test strategy, 194-195
training plans, 196
metadata management, 494-496
capture use cases, 494-495
maintenance workflow use cases, 495
software package installations, 495
stakeholder signoff, 496
user prototypes, 496
project analytics, 410
transactional data cleansing, 135-136
development
application, 389-390
BI SDLC, 252
    analytics components, 253
data integration, 254
data warehouses, 253
final metadata, 254-255
support materials, 252
system testing, 254
testing, 254
testing preparations, 252-253
change management, 516
custom transactional applications, 109, 120-121
architecture, 112-114
business modeling, 110-111
construction phase, 120
design, 115-119
development, 120-121
elaboration phase, 114-115
final approval, 114
inception phase, 109
requirements, 111-112
testing, 119-122
databases
    anomalies, 439
    constraints, 440
data quality-specific tasks, 441-443
data quality
dashboards, 448-450
remediation, 459
data steward tasks, 407-408
ILM renovation projects, 305
data store renovation processes, 305-306
processes, confirming, 306-307
ILM SDLCs, 278-279
legacy claims renovation project
    confirmation, 321-322
data renovation processes, 318-319
data stewardship processes, 320-321
healthcare data retention criteria, 317-318
MDM, 196-200
    CRUD components, 197
data store, 197-198
development testing, 198-200
orchestration components, 197
support materials, 196
user interface components, 197
metadata, 497-498
metadata repositories, 491
teams
    confirming, 107
    performance objectives, 140
dimensional layer logical data models, 410
dimensional structures, 55
direct performance measures, 446
disaster recovery
    BI SDLC, 232
    MDM, 184
    pharmaceutical customer project, 207
discovery (big data), 330-331, 345
data
    ownership, 333
    stewards, 332-333
defined, 619
new sources, analyzing, 345-346
quality control checkpoint, 349-350
security, 348-349
staging for, 334-335
subject areas, 346-348
distribution measures, 447

E
EDM (Enterprise Data Management)
    information governance
        organization integration, 383-384
    operational reporting, 385-386
EIM (Enterprise Information Management)
    big data. See big data BI. See BI
data management
    data modeling, 9
defined, 9
    information governance relationship, 10
defined, 7
functions, 8
information governance relationship, 35-36
interconnections, 62
MDM. See MDM
privacy and security
information governance integration, 33-34
relationships, 35-36
transaction processing. See transaction processing projects
elaboration phase (construction iteration), 114-115
enforcement policies, 396-397
standards, 407
enterprise change management challenge recommendations, 510
Enterprise Information Management. See EIM
enterprise organization model, 40
execution
ILM renovation projects, 307
post-renovation assessment, 309-310
processes, 307-308
legacy claims renovation project, 322
data stewards, instantiating, 322-323
ongoing data stewardship processes, 323-324
post renovation assessment, 325-326
renovation processes, 324-325
executive change management challenges, 508
external compliance reviews, 375
extract processes, 58

F
failures, 2-3
finance and operations processes BI, 52
transaction processing, 42-43
finance consolidation project, 149-150, 564
blueprint breakdown, 152-157
planning, 150-151
SDLC activities and tasks, listing of, 564
forecasting, 54
foreign key constraints, 440
forms
COTS transactional applications, 146
functional specifications, 134
functional focused data stewards, 16
functional grouping model, 41

HIPAA compliance project, 397-398
healthcare company BI project, 258
analytics requirements, 260-261
data integration logical, 266-269
physical, 269
data warehouse design, 263-265
information governance considerations, 259
logical design, 262-263
SDLC tasks and activities, listing of, 598
healthcare data quality ICD-10 compliance project, 461
H00-H59 criteria, confirming, 463
ICD-10 H00-H59 code extensions review, 463
ICD-10 master data, 461
remediation build, 465-466
deployment, 466
design, 464-465
scope, 462-463
healthcare legacy claims data storage project. See legacy claims data storage project
healthcare metadata project, 499
analytics package, 500
business/technical metadata, capturing, 500-501
claims captures, 499-500
data warehouse data quality criteria, 500
navigational metadata, capturing, 501
reporting, 501-502
high-profile/high-risk business processes, prototyping, 135
HIPAA compliance project, 397-398
hubs (MDM), 164-165
ICU patient monitoring, 341

IGC (information governance council), 11
defining, 72-73

EDM
integration, 383-384
operational reporting, 385-386

HIPAA compliance project, 397-398
leaders, 373-374
costs/time benefits, monetizing, 373
ongoing activities, reviewing, 373-374
model, 372
operational responsibilities
data quality, 377-379
data stewardship, 376-377
metadata management, 379-381
security, 381-382
roles and responsibilities, 78-79
stakeholders, 374-375

ILM (information life cycle management), 34-35, 271
audits, 420-421
BI, 230
COTS transactional applications, 147
big data, 61
data definition, 271
data ownership, 272
objectives, 279-280
requirements, 273
data retention
auditing process, 282
criteria, defining, 285-286
defined, 272
packages, installing, 289
requirements, 273
technologies, 276-277, 288-289
data steward task, 406
data warehouses/data marts, 57
defined, 621
define phase, 284
data retention criteria, 285-286
data retention technology review, 288-289
data store landscape review, 286-287
implementation plan development, 289-291
development SDLCs, 278-279
EIM environments, 273-275
implementation, 291
data stewards, 292-295
production cutover tasks, 291-292
information governance, 34-35
monitoring, 295
data ownership processes, 297-298
data stewardship, 298
policies and standards, 295-297
regulatory compliance, 299-300
technologies, 298-299
policies
defining, 281-282
existing, identifying, 280-281
legacy claims renovation, 326-327
monitoring/controlling, 295-297
requirements, 282
validating, 283-284
proactive, 278
renovation projects, 278, 300
activities and tasks, listing of, 613
compliance and cost assessments, 304
data retention problems, identifying, 301-303
data store problems, reviewing, 303
data store renovation processes, 305-306
development, 305
execution, 307
legacy claims. See legacy claims renovation project
post-renovation assessment, 309-310
process confirmation, 306-307
processes, performing, 307-308
renovation plans, 304
SDLC activities and tasks, listing of
healthcare company legacy claims renovation project, 614
proactive, 611
renovation projects, 613
strategies, 279
data retention, 280
draft policies and compliance documents, 281-282
existing policies, identifying, 280-281
ILM process high-level objectives, 279-280
validating policies and compliance documents, 283-284
immediacy (big data), 339-340
implementation
BI SDLC, 257
ILM, 291
data stewards, instantiating, 292-295
production cutover tasks, 291-292
MDM, 177
analysis, 177
change management impact assessment, 186
CRUD components, 189, 197
data integration, 182-184
data load components, 192-194
data store development, 197-198
data store model, 190-192
deployment, 200-201
designs, 186-187
development, 196-200
functionality prototypes, 184-185
hub user interface designs, creating, 187-188
nonfunctional requirements, 185
orchestration components, 188, 197
package configuration, 195
requirements gathering. See MDM, requirements support materials, 196
testing, 194-195
user interface components, 197
metadata
assessment, 492
deployment, 498-499
design, 494-496
development, 497-498
future-state requirements, 492
governance process documents, 493
organizational model, 493
solution outline, 492-494
technology stacks, 493-494
improve phase (data quality), 452
programs, 452-453
remediation, 453-454
inception phase (construction iteration), 109
incorrect data anomaly, 439
indirect performance measures, 446
information governance components, 10
council. See IGC
data management relationship, 10
defined, 10, 621
EIM relationship, 35-36
issues RASIC chart, 88
MDM interrelationship, 161-162
information life cycle management. See ILM
information security audits, 92
initial staging landing zones, 58
initiate phase
agile SDLCs, 105
construction iterations planning, 108
organizational requirements, 107-108
requirements, 105-106
solution architecture, 106-107
custom transactional applications, 126-128
installing data retention packages, 289
integrated big data, 621
internal compliance reviews, 375
invalid data anomalies, 440
issues
management, 375
resolution, 413-415
IT and business project maintenance RASIC chart, 86
K
knowledge sharing sessions, 375
KPMs (Key Performance Measures)
BI, 228
data steward task, 409
healthcare providers example, 499-500
pharmaceutical company sales/marketing project, 425-426
L
landing zones, 58-59
lead data stewards, 403-404
leaders (IGC), 373-374
costs/time benefits, monetizing, 373
ongoing activities, reviewing, 373-374
legacy claims renovation project, 311
activities and tasks, listing of, 614
assessment
existing policies/compliance documents, 315-316
opportunities, identifying, 313-314
policies/compliance documents, validating, 316-317
technical deep dive, 314-315
development
confirmation, 321-322
data renovation processes, 318-319
data stewardship processes, 320-321
healthcare data retention criteria, 317-318
execution, 322
data stewards, instantiating, 322-323
ongoing data stewardship processes, 323-324
post renovation assessment, 325-326
renovation processes, 324-325
ILM policies and standards, 326-327
information governance considerations, 311
monitoring, 326-327
life cycles
big data, 343-344
captures, 350-352
clinical trials project. See clinical trials project control, 350
data definition tasks, 352-354
discovery, 345
new data sources, analyzing, 345-346
new integrated data, identifying, 357-359
opportunities, reviewing, 355-357
prototyping, 361-362
quality control, 354-355
quality control checkpoint, 349-350
security, 348-349, 360
subject areas, 346-348
usage phase, 355
use cases, developing, 359-360
data quality, 434-436
audit phase, 445
criteria, defining, 436-437
dashboard development, 448-450
data integration, 443-445
define phase, 436
development thread tasks, 438-439
improve phase, 452
measurement processes, developing, 446-448
modeling and database design, 439-443
ongoing checkpoint processing, 451
periodic audits, 451
preventive processes, 437-438
programs, 452-453
remediation, 453-454
scope, defining, 436
metadata
capture phase, 481-483
maintenance, 487-488
usage patterns, 484-487
LOB (line of business), 20, 41
logical data models
creating, 112-113
data steward task, 406
dimensional layer, 410
logical design
BI, 234
analytics, 234-235
data integration, 238-240
data warehouses, 236-238
testing, 241-243
healthcare company BI project
analytics, 262-263
data integration, 266-269
data warehouses, 263-265
loyalty program analytics, 342
COTS transactional applications
conversion support, 145
policies, 145
defined, 48
locations, mapping, 137
management. See MDM
MDM (master data management)
analysis
data integration, 182-184
functionality prototypes, 184-185
nonfunctional requirements, 185
requirements. See MDM, requirements, 177-182
architecture, 49, 173-176
component layers, 174
components, 163-165
defined, 621
gap analysis, 174-175
optimal model, selecting, 174
solution/product selection, 175-176
capabilities maturity model assessment, 169
CASE tools, 478
conceptual data integration models, 184
CRUD services, 621
defined, 8
deployment, 200-201
design, 186-187
CRUD components, 189
data load components, 192-194
data store model, 190-192
hub user interface designs, creating, 187-188
orchestration components design, 188
package configuration, 195
testing, 194-195
training plans, 196
development
CRUD components, 197
Index 639

regulatory compliance, 299-300
technologies, 298-299
legacy claims renovation project, 326-327
policies, 396

N
navigational metadata, 27-28, 475-476
BI SDLC, 233
capturing, 501
healthcare company BI project, 268-269
near-line storage, 277
new report specifications, 134
nonfunctional requirements (MDM), 185

O
one-time initial data loads
healthcare company BI project, 268
logical design, 240
ongoing operations
data quality checkpoint processing, 451
data stewardship, 411
business community integration, 423-424
data ownership administration, 412-413
data quality, 415-417
ILM audits, 420-421
issue-resolution, 413-415
MDM, 417-420
security audits, 421-422
IGC
leaders review, 373-374
stakeholders, 375
information governance integration, 86
data quality audits, 89
data quality remediation, 90
information governance issues, 88
information security audits, 92
IT and business functionality, 86
regulatory and compliance audits, 93
reporting and information audits, 87
reporting, 374
operations, 30
big data use cases, 359
change management challenges
listing of, 509
recommendations, 512
IGC/CDO ongoing responsibilities
data quality, 377-379
data stewardship, 376-377
metadata management, 379-381
security, 381-382
implementations, 50-51
MDM, 50-51
metadata, 30, 477, 478
ongoing. See ongoing operations
policy procedures
adding/changing, 394-395
creating, 391
deploying, 395-396
enforcement, 396-397
monitoring, 396
potential changes, reviewing, 392-394
real-time big data integration, 340-341
reporting, 53
teams
confirming, 108
COTS transactional projects, 140
user metadata, 31
opportunities (big data), 339
immediacy, 339-340
real-time decisions, 340
reviewing, 355-357
orchestration services, 164
defined, 621
design, 188
development, 197
pharmaceutical customer project, 206
orders and production processes, 45-46
Oreo Cookie tweet big data example, 339-340
organizations
BI
integration, 387-388
strategies, 225-227
CDO role, 11-12
challenges
enterprise recommendations, 510
executive, 508
operations, 509, 512
project level recommendations, 513
projects, 509
transition, 507
change management
analysis, 514-515
deployment, 517
design, 515-516
development, 516
importance, 505-507
COTS applications, 139-141
change management, 139-141
development team
performance objectives, 140
end-user security roles, 141
information governance organization requirements, 141
operations/maintenance teams, 140
stakeholder management, 139
data stewards, 12-13
defined, 11, 620
development teams,
confirming, 107
expanding roles, 383
IGC, 11
ILM policies, monitoring/controlling, 296
information governance
  application development integration, 389-390
  CDO role, 11-12
  data stewards, 12-13
  defined, 11
  EDM integration, 383-386
  expanding roles, 383
  HIPAA compliance project, 397-398
  IGC, 11
  MDM integration, 386
  reporting models, 13-14
  support requirements, 108
  maintenance teams, 108
MDM
  change management
    impact assessment, 186
    hierarchy requirements, 179
  ongoing, defining, 172-173
  operations, 379-381
  data quality, 377-379
  data stewardship
    community, 376-377
    metadata management, 379-381
  security, 381-382
  teams, confirming, 108
policy procedures
  adding/changing, 394-395
  creating, 391
  deploying, 395-396
  enforcement, 396-397
  monitoring, 396
  potential changes, reviewing, 392-394
  reporting models, 13-14
  requirements, 407
setting up. See setting up
  information governance organizations
  skills assessment, 351

skills gap analysis
  BI, 226
  MDM, 173
transaction processes, 40-41

ownership (data)
  analytics, 54
  BI, 212
  big data, 61, 333-334
    determining, 352
    overview, 333
    rules, applying, 334
  data integration, 59
  data stewardship
    administration, 412-413
  data warehouses/data marts, 57
    defined, 271
  EIM environments, 273-275
  objectives, 279-280
  owners, 620
  processes, monitoring, 297-298
  requirements, 273

P
  parameterized reporting, 53
  party domain requirements, 177-178

performance
  data quality measures, 446
  KPMs. See KPMs
  BI, 228
  data steward task, 409
  healthcare providers
    example, 499-500
  pharmaceutical company
    sales/marketing project, 425-426
  periodic data quality audits, 451
  periodic planning, 375
  pharmaceutical company
    case study, 624-625
    clinical research big data project WBS, 617-616
    clinical trials. See clinical trials project
    customer project, 201-202
  activities and tasks, listing of, 580
  business services, 206
  customer requirements, 205
  data integration model, 207
  data model, 206
  data stewardship, 205
  disaster recovery/business continuity, 207
  information governance considerations, 203-204
  organization change readiness, 207
  SDLC activities and tasks, listing of, 580
  security use cases, 206
  source data quality, 207
  source systems, 206
  user interfaces, 206
  finance consolidation project, 564
  blueprint breakdown, 152-157
  planning, 150-151
  SDLC activities and tasks, listing of, 564
  sales/marketing KPMs project, 425-426
  physical data models
    BI
    analytics design, 244-245
    data integration, 249-252
    healthcare company project, 269
    warehouses, 246-249
    custom transactional applications, 116
    data steward task, 406
    technical, 247-248
  policies, 76
    adding/changing, 394-395
    change reviews, 374-375
    COTS
    master data, 145
    transactional applications, 134
creating, 391
data quality, 76
deploying, 395-396
enforcement, 396-397
HIPAA compliance project, 398
ILM
defining, 281-282
existing, identifying, 280-281
legacy claims renovation project, 315-317, 326-327
monitoring/controlling, 295-297
requirements, 282
validating, 283-284
information governance, 73
assurance, 74
defining, 76
direction, 74
examples, 74-75
guidelines, 73
requirements, 73
standards, 73-74
MDM, 76
metadata management, 76
monitoring, 396
potential changes, reviewing, 392-394
privacy/security, 76
publishing, 76
standards, 390
populating metadata, 480
post-renovation assessments, 309-310
precise technical data quality
dimension, 21, 432
predictive modeling, 54
preventive data quality processes, 437-438
primary key constraints, 440
privacy. See security
proactive ILM, 278, 611
proactive metadata reuse, 486
problems
data retention, identifying, 301-303
data stores, reviewing, 303
processes
baseline
business, reviewing, 131
current, 131
business, developing/leveraging, 133
buying (COTS), 129-130
COTS transactional applications, confirming, 144-147
data
ingress, 58
quality, 58, 434
renovation, 318-319
retention auditing, 282
stewardship, 320-321
dependency diagrams, 133
extract, 58
finance and operations
BI, 52
transaction processing, 42-43
high-profile/high-risk business, prototyping, 135
ILM
data ownership, 297-298
data stewardship, 298
ILM renovation
confirming, 306-307
data store renovation, 305-306
performing, 307-308
layers, 58-59
legacy claims renovation, 324-325
load, 59
orders and production, 45-46
preventive data quality, 437-438
publish, 59
sales and marketing
BI, 52
transaction processing, 43-45
subscribe, 58
transaction. See also transaction processing projects
defined, 8
finance and operations, 42-43
orders and production, 45-46
organization models, 40-41
sales and marketing, 43-45
transactional system development, 102
weak versus strong matrix business models, 41
transform, 58
product domain requirements, 179
production processes, 45-46
programs
data quality, 452-453
MDM, developing, 176
projects
change management challenges
listing of, 509
recommendations, 513
data steward tasks
BI, 409-410
common, 405-407
MDM, 409
integration overview, 81-82
RASIC chart, 83
prototypes
big data solutions, 361-362
core data integration logic, 251
COTS transactional applications
forms, 146
reports, 146-147
high-profile/high-risk business processes, 135
MDM functionality, 184-185
metadata users, 496
project analytics, 410
user interfaces, 115-116, 407-408
provider costs
analytics (healthcare project), 263
logical dimensional data model, 264
publishing, 480
metadata, 480
policies, 76
processes, 59
push versus pull metadata, 486

Q
quality (data)
analysts, 16, 54, 81
Atomic Healthcare data warehouse, 265
audits, 89, 416-417
bad, 18
BI, 212, 231
big data, 61
checkpoint, 349-350
control checkpoint, 354-355
criteria, 353
discovery, 336
integration, 337
streaming, 338
business dimension, 20
checkpoints, 22-23
clinical trials project, 365
constraints, validating, 248-249
criteria
defined, 620
identifying, 410
data integration, 59
data warehouses/data marts, 57
defined, 17, 430
framework, 19, 430
business-process dimension, 433-434
criteria, 431-432
processes, 434
technical dimension, 432-433
healthcare data warehouse criteria, 500
healthcare ICD-10 compliance project, 461
H00-H59 criteria, confirming, 463
ICD-10 H00-H59 code extensions review, 463
ICD-10 master data, 461 remediation, 464-466
scope, 462-463
key data quality elements, 19-20
levels, 18
life cycle, 434-436
audit phase, 445
criteria, defining, 436-437
dashboard development, 448-450
data integration, 443-445
define phase, 436
development thread tasks, 438-439
improve phase, 452
measurement processes, developing, 446-448
modeling and design, 439-443
ongoing checkpoint processing, 451
periodic audits, 451
preventive processes, 437-438
programs, 452-453
remediation, 453-454
scope, defining, 436
logical data warehouses, 237
management
bad data quality, 18
business dimension, 20
checkpoints, 22-23
data quality levels, 18
defined, 17, 620
framework, 19
key data quality elements, 19-20
processes dimension, 22
process types, 23
technical dimensions, 21-22

MDM sources, 183
measures, 447
organization operations, 377-379
policies, 76
processes, 58
dimension, 22
types, 23
remediation
criteria, defining, 456
critical data elements, identifying, 456
deployment, 460-461
design, 457-459
development, 459
overview, 453
RASIC chart, 90
scope, 454-455
SWAT projects, 454
teams, 453
reporting users, 447
stewardship management, 415-417
audit scenario, 416-417
issue resolution scenario, 415-416
technical dimensions, 21-22
queries (ad hoc), 53

R
RASIC (Responsibility, Approves, Supports, Informed, Consults) charts
data quality audits, 89
remediation, 90
data stewardship, 77
defined, 620-621
information governance issues, 88
project integration, 83
IT and business functionality, 86
regulatory and compliance audits, 93
reporting and information audits, 87
security audits, 92
reactive ILM SDLC tasks and activities, listing of, 613
registry architectural style, 49
regulations
  audits, 93
  ILM compliance, 299-300
  reporting, 294-295
release and deployment tasks, 122
remediation (data quality), 453-454
  criteria, defining, 456
  critical data elements, identifying, 456
  deployment, 460-461
design, 457-459
development, 459
healthcare ICD-10 compliance project, 464-466
overview, 453
scope, 454-455
SWAT projects, 454
teams, 453
renovation projects (ILM), 278, 300
  compliance and cost assessments, 304
  data retention problems, identifying, 301-303
data store problems, reviewing, 303
development, 305
data store renovation processes, 305-306
  processes, confirming, 306-307
execution, 307
  post-renovation assessment, 309-310
  processes, performing, 307-308
legacy claims, 311
  confirmation, 321-322
data renovation processes, 318-319
data stewards, instantiating, 322-323
data stewardship
  processes, 320-321
  execution, 322
  existing policies/compliance documents, 315-316
healthcare data retention criteria, 317-318
ILM policies and standards, 326-327
information governance considerations, 311
monitoring, 326
ongoing data stewardship processes, 323-324
  opportunities, identifying, 313-314
  policies/compliance documents, validating, 316-317
post renovation assessment, 325-326
renovation processes, performing, 324-325
  technical deep dive, 314-315
plans, developing, 304
reporting audits
  IGC leaders, 373
  RAS IC chart, 87
compliance, 374
COTS transactional applications, 146-147
data quality, 377-379, 447
deletion, 308
EDM operational, 385-386
healthcare metadata project, 501-502
information governance models, 13-14
metadata, 380-381, 487
new report specifications, 134
ongoing operational, 374
  operational, 53
  parameterized, 53
  regulatory, 294-295
security, 382
usage creation, 374
value creation, 374
repositories (metadata), 488
  administration, 489-491
  administrator role, 490-491
  configuration management, 490
technical/functional tasks, 489-490
building versus buying, 491
captures, 481-483
user requirements, 488-489
requirements accounts, 178
BI
  analysis, 228
  data integration, 230-233
  data warehouses, 228-230
  information life cycles, 230
big data
  functional, 359
  retention, 354
  security, 353-354, 360
  users, 355
business services, 181
conceptual data models, 182
contracts, 178
custom transactional applications
  construction iteration, 111-112
  initiate phase, 105-106
  organizational, 107-108
data
  integration, 182-184
  ownership, 273
  retention, 273
  stewardship, 179-180
healthcare company BI project, 260-261
ILM policies, 282
information governance policies, 73
MDM implementations, 177-182
  account/contracts, 178
business services, 181
conceptual data models, 182
data integration, 182-184
data stewardship, 179-180
information security use cases, 182
nonfunctional, 185
organizational hierarchy, 179
party domain, 177-178
product domain, 179
user interfaces, 181
metadata repository user, 488-489
nonfunctional, 185
organizational confirming, 407
hierarchy, 179
party domain, 177-178
product domain, 179
retention, 117
security identifying, 112
use cases, 182
user interfaces, 181
Responsibility, Approves, Supports, Informed, Consults.
See RASIC charts
retention (data)
agile transactional application data, 117
auditing process, 282
big data
discovery, 336
integration, 337
requirements, 354
streaming, 338
criteria, defining, 285-286
defined, 272, 620
EIM environments, 273-275
healthcare criteria, 317-318
objectives, 280
packages, installing, 289
problems, identifying, 301-303
requirements, 273
technologies, 276-277, 288-289
rules (business), capturing, 110-111
sales and marketing processes
BI, 52
transaction processing, 43-45
sandbox environments, 423-424
SCCB (software configuration control board), 622
scope (data quality), 436
healthcare ICD-10 compliance project, 462-463
remediation, 454-455
scorecards, 54
SDLC (Systems Development Life Cycle), 163
agile method
custom transactional processing activities and tasks, listing of, 543
healthcare claims reengineering project
WBS, 554
information governance, threading, 103
iterative phases, 103
BI
agile, applying, 216-217
analysis phase, 227-228
analytics, 213-214, 234-235, 253
architectural strategy, 223-225
business impact, 233-234
client business/IT environment assessment, 220-221
data integration, 230-233, 254
data stewardship, 217-219
data warehouses, 228-230, 253
deployment, 255-257
development, 252, 254
disaster recovery/business continuity, 232
final metadata development, 254-255
healthcare company project. See healthcare company BI project implementation checklist, 257
logical data integration design, 238-240
logical data warehouse design, 236-238
logical design, 234
ongoing organizational strategies, 225-227
physical analytics design, 244-245
physical data integration, 249-252
physical data warehousing, 246-249
program strategy, 221-223
strategy phase, 220
support materials, 252
system testing, 254
tasks and activities, listing of, 587
technical design, 243
testing, 241-243, 252-253, 255-256
training, 252
transformations, 216
COTS custom transaction processing
baseline business processes, reviewing, 131
baseline current process activity, 131
blueprinting completion, 132-137
blueprint walkthrough, 138
business blueprint phase, 130
index

change management,  139-141
finance consolidation project. See finance consolidation project
gap analysis, 132
mapping and construction phase, 144-148
SDLC activities and tasks, listing of, 561
testing, 142-143
ILM
data ownership processes, monitoring, 297-298
data retention, 280, 285-286, 288-289
data stewardship, 292-295, 298
data store landscape review, 286-287
define phase, 284
defining draft policies and compliance documents, 281-282
existing policies, identifying, 280-281
healthcare company legacy claims renovation project WBS, 614
ILM process high-level objectives, 279-280
implementation, 289-291
monitoring, 295
policies and standards control, 295-297
proactive activities and tasks, listing of, 611
production cutover tasks, 291-292
regulatory compliance, 299-300
renovation projects. See ILM, renovation projects strategies, 279
tasks review, 278-279
technologies, monitoring, 298-299
validating policies and compliance documents, 283-284
MDM. See MDM
transactional processing. See transaction processing projects
security analytics, 54
audits, 92, 421-422
big data, 61, 348-349
discovery, 336
integration, 337
requirements, 353-354, 360
streaming, 338
broader IT security interlocking, 33
CISO role, 34
data steward task, 406
warehouses/data marts, 57
declared, 622
EIM integration, 33-34
end-user roles, 141
organization operational responsibilities, 381-382
overview, 32
pharmaceutical customer project, 206
policies, 76
requirements, identifying, 112
testing, 198-199
BI, 243, 255-256
COTS transactional applications, 142-143
custom transactional applications, 120
MDM, 195, 198-199
use cases, 182
semantic federation, 339
services CRUD. See CRUD services
custom transactional applications, 118
identification, 113-114, 408
MDM requirements, 181
orchestration, 164
defined, 621
design, 188
development, 197
pharmaceutical customer project, 206
setting up information governance organizations
CDO roles, 71-72
data stewardship community, 76-77
IGC, defining, 72-73
ongoing operations integration, 86
data quality audits, 89
data quality remediation, 90
information governance issues, 88
IT and business functionality, 86
regulatory and compliance audits, 93
reporting and information audits, 87
security audits, 92
organization versus program, 70
policies assurance, 74
data quality, 76
defining, 76
direction, 74
terms, 74-75
guidelines, 73
information governance, 73
MDM, 76
metadata management, 76
privacy/security, 76
publishing, 76
requirements, 73
standardization, 74
standards, 73
project integration overview, 81-82
RASIC chart, 83
roles and responsibilities
   data stewardship, 79-81
   IGC, 78-79
structure, defining, 71
shadow IT departments, 423-424
Singapore traffic flow with
   sensor data improvements, 339
   smart meter management, 341
   social networks, 339-340
   software configuration control
board (SCCB), 622
sources
   big data
      access, 356
      analyzing, 345-346
MDM
   data quality, 183, 231
   mapping, 410
   system profiling, 183
pharmaceutical customer
   project, 206-207
   system profiling, 183,
      230-233, 410
   systems, 206
stakeholders
   COTS
      readiness surveys, 144
      transactional applications,
         139
   IGC, 374-375
standards
   adherence, 106-107
   defined, 390
   enforcement, 407
   information governance
   policies, 73
stewardship, 16
   big data projects, 332-333
   captures, 350-352
   control, 350
   data definition tasks,
      352-354
   discovery, 345
   new data sources,
      analyzing, 345-346
   new integrated data,
      identifying, 357-359
   opportunities, reviewing,
      355-357
   prototyping, 361-362
   quality control, 349-350,
      354-355
   security, 348-349, 360
   subject areas, 346-348
   usage phase, 355
   use cases, developing,
      359-360
   BI SDLC, 217-219
   business, 15-16
   business community
      integration, 423-424
   common characteristics, 17
   community structure,
      403-405
defined, 14, 620
   identities, 402
   IGC leaders, 373
ILM
   instantiating, 292-293
   ongoing, 293-295
   policies, monitoring/
      controlling, 296
   processes, monitoring,
      298
information governance
   business, 15-16
   common characteristics,
      17
defined, 14
   organization, setting up,
      76-77
   responsibilities, 14-15
   technical, 16-17
legacy claims renovation
   project, 320-323
   MDM, 80
   requirements, 179-180
   user interface, 188
   metadata, 80-81
   ongoing operations, 411
   data ownership
      administration, 412-413
   data quality management,
      415-417
   ILM audits, 420-421
   issue-resolution, 413-415
   MDM, 417-420
   security audits, 421-422
organization operations,
   376-377
pharmaceutical company
   sales/marketing KPMs
   project, 425-426
pharmaceutical customer
   project, 205
project tasks
   BI, 409-410
   common, 405-407
custom development
   projects, 407-408
   MDM, 409
RASIC charts, 77, 620-621
   roles and responsibilities,
      79-81, 402-403
   analytic, 80
   community, 79-80
data quality analysts, 81
data stewards, 79-80
   MDM stewards, 80
   metadata management,
      80-81
technical, 14-15
typical, 14-15
storage
   cloud, 277
   Hadoop, 277
   near-line/tape, 277
   traditional, 277
   virtual, 277
strategies
   BI, 220
   architectural, 223-225
   client business/IT
      environment assessment,
         220-221
   ongoing organizational,
      225-227
   program, 221-223
ILM, 279
   data retention objectives,
      280
defining draft policies and
   compliance documents,
      281-282
   existing policies,
      identifying, 280-281
ILM process high-level objectives, 279-280
validating policies and compliance documents, 283-284
MDM, 168
client business/IT strategy assessment, 168-169
master data management, 170-171
ongoing organizational definition, 172-173
streaming big data, 338
structural metadata, 26-27, 473-474
subscribe processes, 58
SWAT remediation projects, 454
systems, testing, 254
Systems Development Life Cycle. See SDLC
metadata stacks, 493-494
users, 31, 478
testing BI, 241-243
development, 254
preparations, 252-253
security, 255-256
system, 254
testing cycles, 255-256
types, 241-242
user acceptance, 243
COTS transactional applications, 142-143
components, 147-148
project metadata, capturing, 143
security test cases, 142-143
strategies, 142
custom transactional applications
construction phase, 121-122
elaboration phase, 119-120
strategies, defining, 119
transition, 122
data quality remediation components, 459
information security test cases, defining, 120
MDM development, 198-200
implementation, 194-195
unit, 245
user acceptance, 255-256
timely technical data quality dimension, 21, 432
traffic flow with sensor data improvements, 340
transactional architectural style, 49
transactional metadata, 25
transaction processing projects
core business model processes, 42
finance and operations, 42-43
orders and production, 45-46
sales and marketing, 43-45
COTS, 129
baseline business processes, reviewing, 131
baseline current process activity, 131
blueprinting completion, 132-137
blueprint walkthrough, 138
business blueprint phase, 130
change management, 139-141
finance consolidation project. See finance consolidation project gap analysis, 132
mapping and construction phase, 144-148
pharmaceutical company finance consolidation project WBS, 564
processes, buying, 129-130
SDLC tasks and activities, listing of, 561
testing, 142-143
transactional data cleansing designs, 135-136
custom, 104
agile SDLC activities and tasks. listing of, 543
architecture, 112-114
business modeling, 110-111
claims reengineering project. See claims, reengineering project configuration and change management tasks, 123-124

T

tape storage, 277
technical BI design, 243
analytics, 244-245
data integration, 249-252
data warehousing, 246-249
training, 252
data quality
checkpoints, 23
dimension, 21-22, 432-433
data steward, 16-17
logical data warehouses, 237-238
metadata
capturing, 500-501
logical data warehouses, 237-238
support, 486
technologies
data retention, 276-277, 288-289
ILM, monitoring, 298-299

 finance and operations, 42-43
 orders and production, 45-46
 sales and marketing, 43-45
 COTS, 129
 baseline business processes, reviewing, 131
 baseline current process activity, 131
 blueprinting completion, 132-137
 blueprint walkthrough, 138
 business blueprint phase, 130
 change management, 139-141
 finance consolidation project. See finance consolidation project gap analysis, 132
 mapping and construction phase, 144-148
 pharmaceutical company finance consolidation project WBS, 564
 processes, buying, 129-130
 SDLC tasks and activities, listing of, 561
 testing, 142-143
 transactional data cleansing designs, 135-136
 custom, 104
 agile SDLC activities and tasks. listing of, 543
 architecture, 112-114
 business modeling, 110-111
 claims reengineering project. See claims, reengineering project configuration and change management tasks, 123-124

 T

tape storage, 277
technical
BI design, 243
analytics, 244-245
data integration, 249-252
data warehousing, 246-249
training, 252
data quality
checkpoints, 23
dimension, 21-22, 432-433
data steward, 16-17
logical data warehouses, 237-238
metadata
capturing, 500-501
logical data warehouses, 237-238
support, 486
technologies
data retention, 276-277, 288-289
ILM, monitoring, 298-299

 finance and operations, 42-43
 orders and production, 45-46
 sales and marketing, 43-45
 COTS, 129
 baseline business processes, reviewing, 131
 baseline current process activity, 131
 blueprinting completion, 132-137
 blueprint walkthrough, 138
 business blueprint phase, 130
 change management, 139-141
 finance consolidation project. See finance consolidation project gap analysis, 132
 mapping and construction phase, 144-148
 pharmaceutical company finance consolidation project WBS, 564
 processes, buying, 129-130
 SDLC tasks and activities, listing of, 561
 testing, 142-143
 transactional data cleansing designs, 135-136
 custom, 104
 agile SDLC activities and tasks. listing of, 543
 architecture, 112-114
 business modeling, 110-111
 claims reengineering project. See claims, reengineering project configuration and change management tasks, 123-124
construction iteration, 108-109, 111-112, 120
data steward tasks, 407-408
design, 115-119
development, 120-121
elaboration phase, 114-115
final approval signoff, 114
inception phase, 109
initiate phase, 105
initiation requirements activity, 105-106
organizational requirements, 107-108
release and deployment tasks, 122
solution architecture, 106-107
testing, 119-122
transition iteration, 122
transition testing, 122
data ownership/retention, 274
defined, 8, 622
healthcare claims reengineering project WBS, 554
organization models, 40-41
transactional system development, 102
weak versus strong matrix business models, 41
transformations, 215-216
transform processes, 58
transition change management challenges, 507
transition iteration (custom transactional applications), 122
configuration and change management tasks, 123-124
release and deployment tasks, 122
testing, 122
Twitter medication message data (clinical trials project), 364-365
unique key constraints, 440
unique technical data quality dimension, 21, 432
unit testing, 245
updating common vocabulary, 111
usage creation reporting, 374
usage phase
big data, 355
clinical trials project, 366
new integrated data, identifying, 357-359
opportunities, reviewing, 355-357
prototyping, 361-362
security, 360
use cases, 359-360
metadata, 484-487
ongoing business and technical, 486
proactive reuse, 486
push versus pull, 486
reporting, 487
use case analysis, 408
user interfaces
MDM
data stewardship, 188
development, 197
hub user interface designs, creating, 187-188
requirements, 181
pharmaceutical customer project, 206
prototyping, 115-116, 408
users, 30-31
acceptance testing, 243
BI, 255-256
MDM, 199
big data requirements, 355
data quality reporting, 447
MDM test strategy, 194-195
metadata, 30-31, 478
prototypes, 496
repository requirements, 488-489
security roles, 141
valid technical data quality dimension, 21, 432
value creation reporting, 374
virtual storage, 277
vocabulary
capturing, 105-106
updating (construction iteration), 111
W - Z
WBS
BI
healthcare company project SDLC, 598
SDLC tasks and activities, listing of, 587
big data pharmaceutical company clinical research project, 617-616
COTS
custom transactional processing SDLC, 561
pharmaceutical company finance consolidation project, 564
custom transactional processing project, 543
healthcare company claims reengineering project, 554
ILM SDLCs
healthcare company legacy claims renovation project, 614
proactive, 611
renovation project, 613
MDM
pharmaceutical customer project, 580
SDLC activities and tasks, listing of, 579
yearly planning, 375
Zachman best practice, 132